ELECTRICITY AND MAGNETISM

CONSULTING EDITOR

Professor J M Cassels, M A , Ph D , F R S

Lyon Jones Professor of Physics

University of Liverpool

ELECTRICITY
AND
MAGNETISM

by

W. J. DUFFIN
Senior Lecturer in Physics
University of Hull

McGRAW-HILL PUBLISHING COMPANY LIMITED
LONDON
New York · Toronto · Sydney

Published by
McGRAW-HILL PUBLISHING COMPANY LIMITED
McGRAW-HILL HOUSE, MAIDENHEAD, BERKSHIRE, ENGLAND

94008

THIS BOOK HAS BEEN SET IN MONOPHOTO TIMES NEW ROMAN 10/12 PT
AND PRINTED AND BOUND IN GREAT BRITAIN BY
WILLIAM CLOWES AND SONS, LIMITED, LONDON AND BECCLES

PREFACE

Of all the branches of classical physics studied in schools, electricity and magnetism is in my experience the least appreciated as a coherent body of knowledge. For this reason, I have tried to provide in this volume an account of the subject which shows clearly the experimental bases, the development of concepts and the formulation of general laws. At the beginning I assume no more than a sketchy knowledge of school work in both physics and mathematics, and some topics in the latter which may be unfamiliar are covered in appendices to the earlier chapters. Fewer concessions in this direction are made as the book progresses and the final standard is one which should be adequate for Part I (the first two years) of an honours degree course in physics, for most of a general degree course including physics and for the early stages of electrical engineering courses. I also hope that the book may be of interest to sixth formers intending to proceed to university or technical college courses.

The general arrangement of material follows fairly orthodox lines except that in the first chapter I have given rather more attention than is usual to the establishment of the basic concepts of charge and current. Chapters 2–5 deal with electric forces and fields and chapters 7–9 with the corresponding magnetic phenomena. In these chapters, as in most, I have included material which is relevant to other studies, particularly atomic, nuclear and solid state physics: this accounts for the emphasis on dipoles and the introduction to higher multipoles in chapter 4, for instance. The network theory of chapters 6 and 10 goes rather further than is usual in a book of this standard because of its relevance to electronics (now usually taught at an early stage) and because it does not need any new or difficult concepts.

Chapter 11 is an introduction to charged-particle dynamics, included both to show the modern applications of a classical subject and to illustrate the confirmation of laws established for conduction currents. Chapters 12, 13 and 14 deal with the three important electric and magnetic properties of materials, but there is no reason why much of each should not be studied earlier: the initial sections

of chapter 13, for instance, after chapter 5. Chapter 15 gathers together the laws established previously and generalizes them in Maxwell's equations, thus providing the starting point for a more advanced study of the subject. The final chapter deals with certain aspects of measurement not previously encountered and ends with a discussion of units.

It is customary in prefaces to defend one's point of view about unit systems. I take a pragmatic view of the matter and have chosen the MKSA system in preference to the CGS Gaussian because it is more convenient to use the same units in theoretical, computational and practical work and thus avoid the need to remember, look up or calculate conversion factors. The rationalized system is preferable to the nonrationalized because the units for **H** and **D** are then A/m and C/m^2 respectively rather than 4π times these (a situation which would presumably call for two new names). Many authors and readers prefer CGS units, however, and I have discussed the various alternatives liberally in the text and have provided in chapter 16 both a chapter-by-chapter conversion to Gaussian units and a list of the systems used by other authors.

There is, unfortunately, a widespread impression that the use of MKS units is tied to a method of teaching which starts from uniform electric and magnetic fields, a method not favoured by many teachers. I have tried to show that this association is not essential, although I have relegated the magnetic pole to the status of an auxiliary concept rather than a basic one.

A few omissions from the book should not go unexplained. Thermoelectric and chemical effects have received scanty attention because an adequate treatment is outside the scope of the work. Electronic circuitry is omitted completely as being a highly technical subject needing separate treatment, and only a brief account of electrical machinery is provided for the same reason.

Each chapter starts with a brief outline of its aim and subject-matter in a form suitable for those meeting the material for the first time; and ends, where appropriate, with a summary and discussion of the results obtained, in a form which assumes that the chapter has been read and digested. Problems are meant to be solved as an integral part of the study, and the answers in the early stages are for this reason fairly lengthy. Starred problems are rather more difficult and might be left until a second reading. Equations which are the direct result of definitions are so labelled for emphasis. References to further reading at the ends of chapters, and to original

works within chapters, are given in the form *author* (*date*), and details are given at the end of the book. References to original papers are included largely for the benefit of teachers who may wish to pursue certain topics more deeply.

Acknowledgements

My interest in electromagnetism began in undergraduate days when the lecturer was J. A. Ratcliffe and textbooks were early editions of Harnwell and of Page and Adams. If any signs of their influence appear in the following pages, I gratefully acknowledge them. I am also indebted to my colleagues, Dr. A. Cunliffe, Dr. J. I. Jones and Dr. Ronald Shaw, who have so often stimulated me to think further about points I had long regarded as settled. Dr. Jones, Dr. Shaw and my wife have also read and helpfully criticized many points in the manuscript and much of the book would be the worse but for their help. I also thank Mr. T. J. Harvey and Mr. R. Hollings who undertook the arduous task of checking many of the problems.

For permission to quote representative properties of materials in the tables I thank the sources mentioned in the appropriate places. In addition, the Permanent Magnet Association and the International Nickel Company (Mond) Ltd. have most helpfully provided information about magnetic materials.

In spite of all the help I have had from colleagues, from the publishers and their advisers, and from my family (who have shown great forbearance), there must remain mistakes, obscurities and misprints. I hope readers will not hesitate to point out to me any of these so that future editions may be improved.

W. J. Duffin

CONTENTS

GLOSSARY OF SYMBOLS

The symbols used in this book agree generally with those recommended by the International Organization for Standardization (1961) and by the S.U.N. Commission of the International Union of Pure and Applied Physics (1961). The only important changes made are the use of τ for volume (instead of V), **N** for the Poynting vector (instead of **S**) and I_m for moment of inertia (instead of I) to avoid confusion within one formula. Symbols for units are given in tables 1.2 and 16.1 and are not included here.

$\mathbf{A}$, A	Area, particularly plane areas of plates, coils and of cross-sections
$\mathbf{A}$	In appendices, a general vector
A	Atomic weight
$\mathbf{a}$, a	Acceleration
a	Length, particularly of radii
$\mathbf{B}$, B	Magnetic flux density
$\mathbf{B}$	In appendices, a general vector
B	Susceptance
b	Radius of cylinder or sphere; air damping constant of galvanometer
C	Capacitance; general path in space, open or closed
c	Velocity of light *in vacuo*; torsional constant
c_{ij}	Coefficients of capacitance
$\mathbf{D}$, D	Electric displacement
D	Coefficient of diffusion
d	Distance, e.g. between parallel wires
$\mathbf{E}$, E	Electric field strength
$\mathscr{E}$	Electromotive force
e	Electronic charge
$\mathbf{F}$, F	Force; local electric or magnetic field
$\mathbf{G}$, G	External force
G	Conductance; electromagnetic momentum
g	Acceleration due to gravity; g-factor
$\mathbf{H}$, H	Magnetic field strength

$\mathscr{H}$	Magnetomotive force
h	Planck's constant; height
$\mathbf{I}$, I	Magnetic polarization
$\mathbf{I}$, I	Electric current; I_c, conduction current; I_d, displacement current; I_M, Amperian current; I_s, surface current
I_m	Moment of inertia
$\mathbf{J}$, J	Volume current density; J_c, conduction current density; J_d, displacement current density
$\mathbf{J}_s$, J_s	Surface current density; J_{sc}, conduction surface current density; J_{sM}, Amperian surface current density
K	Kinetic energy
k	Boltzmann's constant; coefficient of coupling; 2π/wavelength
$\mathbf{L}$, L	Angular momentum
L	Self-inductance
$\mathbf{l}$, l	Length, particularly of wire, rod or line
$\mathbf{M}$, M	Magnetization
M	Mutual inductance; molecular weight; mass, particularly of proton
m	Mass, particularly of electron
$\mathbf{m}$, m	Magnetic moment
$\mathbf{N}$, N	Poynting vector
N	Number of conductors, turns, etc.
N_A	Avogadro's number
n	Number per unit volume, length or time; general number
$\mathbf{P}$, P	Electric polarization
P	Pole strength; power; resistance in bridge network
P_{XY}	Thermoelectric power
$\mathbf{p}$, p	Electric dipole moment; linear momentum
p_{ij}	Coefficients of potential
Q	Electric charge; resistance in bridge network
$\hat{Q}$	Quality factor of circuit
q	Quadrupole moment
R	Resistance; R_l, resistance per unit length
R_H	Hall coefficient
$\mathscr{R}$	Reluctance
$\mathbf{r}$, r	Polar co-ordinate; distance from point; internal or leakage resistance
$\mathbf{S}$, S	Surface area, closed or open
S_I	Current sensitivity; S_Q, quantity or charge sensitivity; S_V, potential sensitivity

$\mathbf{s}$, s	General displacement or direction in space
$\mathbf{T}$, T_θ	Torque or couple
T	Period of time, particularly of oscillations; absolute temperature
t	Time
U	Potential energy
U_{E}	Electric energy; U_{M}, magnetic energy; U_{EM}, electromagnetic energy
$\mathbf{V}$, V	Potential difference
V_{m}	Magnetic potential difference
$\mathbf{v}$, v	Velocity
W	Work
X	Reactance; unknown resistance
x	Cartesian co-ordinate; distance
$\mathbf{Y}$, Y	Admittance
y	Cartesian co-ordinate
$\mathbf{Z}$, Z	Impedance; $\mathbf{Z}_{\mathrm{k}}$, characteristic impedance; $\mathbf{Z}_{\mathrm{L}}$, load impedance; $\mathbf{Z}_{\mathrm{in}}$, input impedance; $\mathbf{Z}_{\mathrm{m}}$, mutual impedance; $\mathbf{Z}_{\mathrm{p}}$, primary impedance; $\mathbf{Z}_{\mathrm{s}}$, secondary impedance
Z	Atomic number
z	Cartesian co-ordinate
α	General angle; phase angle; temperature coefficient of resistance; attenuation and decay constant; polarizability
α_{m}	Magnetic polarizability
β	General angle; phase angle; second temperature coefficient of resistance; phase constant
γ	Propagation constant; gyromagnetic ratio
Δ	Decrement
δ	Finite difference in; argument of complex impedance
ε_0	'Permittivity of free space'
ε_{r}	Relative permittivity (dielectric constant)
η	Coefficient of viscosity
θ	Polar co-ordinate; angle, particularly of deflection
Λ	Logarithmic decrement
λ	Line charge density; mean free path; wave-length
μ	Mobility
μ_0	'Permeability of free space'
μ_{r}	Relative permeability
μ_{B}	Bohr magneton
ν	Frequency; ν_{c}, cyclotron frequency; ν_{L}, Larmor frequency
π_{XY}	Peltier coefficient

ρ	Volume charge density; resistivity; reflection coefficient in lines
σ	Surface charge density; conductivity; σ_P, polarization charge density
σ_T	Thomson coefficient
τ	Volume; time constant or relaxation time
Φ	Magnetic flux
ϕ	Polar co-ordinate
χ_e	Electric susceptibility
χ_m	Magnetic susceptibility; χ_ρ, mass susceptibility; χ_A, molar susceptibility
Ω	Solid angle
ω	Angular velocity; angular frequency; ω_N, natural angular frequency; ω_0, undamped natural angular frequency, current resonant angular frequency; ω_Q, charge resonant angular frequency; ω_P, parallel *LCR* resonant angular frequency.

PREFIXES USED TO INDICATE MULTIPLES OR SUBMULTIPLES

T	tera	10^{12}	m	milli	10^{-3}
G	giga	10^{9}	μ	micro	10^{-6}
M	mega	10^{6}	n	nano	10^{-9}
k	kilo	10^{3}	p	pico	10^{-12}
c	centi	10^{-2}			

PHYSICAL CONSTANTS
(from Cohen *et al*, 1955)

Electronic charge, $e = 1{\cdot}60206 \times 10^{-19}$ C
Electron rest mass, $m = 9{\cdot}1083 \times 10^{-31}$ kg
Proton rest mass, $M = 1{\cdot}67239 \times 10^{-27}$ kg
$e/m = 1{\cdot}7589 \times 10^{11}$ C/kg; $M/m = 1836{\cdot}12$
Velocity of light, $c = 2{\cdot}99793 \times 10^{8}$ m/s
Avogadro's number, $N_A = 6{\cdot}02486 \times 10^{26}$/kg
Planck's constant, $h = 6{\cdot}62517 \times 10^{-34}$ J-s
Boltzmann's constant, $k = 1{\cdot}38044 \times 10^{-23}$ J/°K

CHAPTER 1

ELECTRIC CHARGE AND CURRENT

Electricity and magnetism, as branches of physics, are principally concerned with the production of and interactions between electric charges and currents. Our first aim must therefore be to establish what is meant by these terms so that we know what we are talking about, and this we shall do by seeing how to recognize charge and current by the effects they produce. These same effects can then be used to provide measures which, though arbitrary, are sufficient to discover experimental laws.

We then examine the relation between current and rate of flow of charge, and agree to define a unit for one in terms of the other: in particular, the ampere is established as the unit of current. Questions about the nature of charge and current arise naturally and are discussed from the atomic aspect in an elementary way.

Finally, we shall introduce and discuss the concepts of the point charge and of charge and current densities which will be needed frequently in later chapters.

1.1 Electric Charge

We recognize a body as *charged* when it attracts nearby light objects, such as small pieces of paper or cork, without touching them. A dry plastic rod, for instance, can be charged by rubbing it on a coat sleeve, or a glass rod by rubbing it with silk, and careful investigation shows that the coat sleeve and the silk are also charged. Friction between two surfaces invariably leaves both of them charged, and even mere contact and separation, as with a vehicle tyre and the road, will usually have the same effect.

The charged condition of one body can be transferred to another which is initially uncharged either by contact between the two or by connecting them together with a metallic wire. The idea thus arises of some entity which is transferred, distinct from the material of the body itself and giving it the property of attracting other bodies. This we call *electric charge*. Materials through which

charge moves freely, like the above metallic wire, are known as *electric conductors* and are usually distinguished from non-conductors or *insulators*, though the difference is often one of degree rather than of kind, as we shall see in chapters 6 and 12.

When the forces between two charged bodies are examined we find that repulsion as well as attraction occurs and that two types of charge can be distinguished. If the type of charge found on the plastic rod above is denoted by R (for resinous, its original name) and that on the glass rod by V (for vitreous), it is found that

V repels V but attracts R
R repels R but attracts V

Any other charged body is found either to repel V and attract R, and is thus indistinguishable from V, or to repel R and attract V and is thus indistinguishable from R. Hence all charges are either resinous or vitreous, and obey the general law that *like charges repel and unlike charges attract.*

We now require a method of detecting charge which is more systematic and sensitive than the above test of attraction for light objects. If possible, the method should also *measure* charge since only then can quantitative laws be discovered. Such a method is provided by the electroscope, consisting of a light pivoted conducting leaf or needle connected to a rigid conducting rod and cap, all mounted in but insulated from a case connected to the earth (Fig. 1.1). The cap consists of a metal container with a small opening in the top.

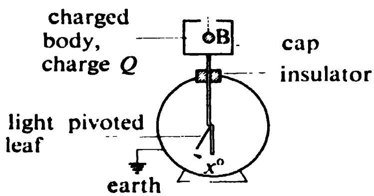
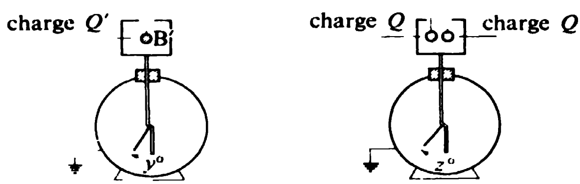


Fig. 1.1. An electroscope.

A charged body B, completely within the cap, produces a deflection of the leaf which does not depend on the position of B and remains even when B touches the inside of the cap. (In elementary terms, the deflection is produced by repulsion between like charges on leaf and rod arising from electrostatic induction (below), but for

a fuller explanation see section 5.8.) If B is of conducting material, touching the inside of the cap leaves a permanent charge on the electroscope and discharges B.

Measurement of charge proceeds in the following way. Two charges are said to be equal in magnitude if they give equal deflections when placed separately in the cap of the electroscope (Fig. 1.2a). If two equal charges are now placed together in the cap, we find no deflection at all if they are unlike charges but an increased



(*a*) (*b*)

Fig. 1.2. ***Measurement of charge:*** Q ***in Fig. 1.1*** $= Q'$ ***if*** $x = y$. ***The deflection*** z ***is not*** $2x$, ***but represents a single charge*** $2Q$.

deflection if they are like charges. We define the magnitude of a single charge to be $2Q$ if it produces the same deflection by itself that two charges of magnitude Q give together (Fig. 1.2b). Similarly, if we can find two charges of equal magnitude which together produce the same deflection as Q, the magnitude of each is $\frac{1}{2}Q$. The electroscope can be calibrated accordingly in terms of an arbitrary unit Q. Extension of the process provides a means of measuring charge over a limited range: the electroscope has become an electrometer.

Experiments with the calibrated electroscope show that when two unlike charges are placed in the cap without contact, the resultant deflection corresponds to the *difference* between them, and that for any number of charges the resultant magnitude is obtained by treating one type as *positive* and the other as *negative*. It is a universal convention that the vitreous charge shall be the positive one.

Conservation of Charge. Superficially similar to but quite distinct from the experiment just described is one in which two charged bodies placed in the cap of an electroscope are brought into contact.

Even if they are conductors, when some redistribution can be expected, there is no change in the total charge indicated by the electroscope Repetition of this experiment with a number of charges leads us to think that *in any isolated system the algebraic sum of the charges is constant* This law, that of the conservation of charge, is confirmed by other facts such as the equality but opposite sign of the charges produced on a glass rod and on a piece of silk when the two are rubbed together the total charge is zero before and after the rubbing Although these experiments are relatively crude there are good reasons for thinking that the law holds exactly (see end of section 1 5)

Electrostatic Induction An important phenomenon can be demonstrated by bringing a charged body B near to an uncharged piece of metal in two detachable parts as in Fig 1 3 If the contact

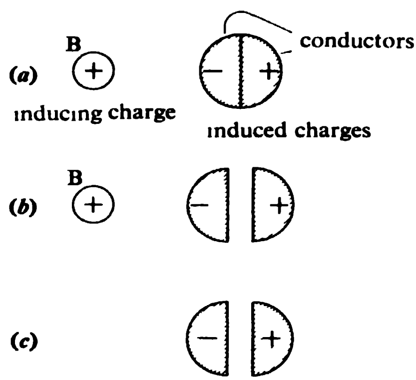


Fig. 1.3. Induced charges in conductors. (a) Electrostatic induction, (b) separation of conductors while* B *is still present; (c) removal of* B *leaves two charged conductors.

between the two parts is broken while B is kept in position they are found to possess equal and opposite charges as shown This separation of charge taking place under the influence of a nearby charge is called *electrostatic induction*

All conductors behave in the same way and we can infer that they must contain charges free to move under the attractive or repulsive force of the inducing charge, though it is not possible at this stage to tell whether these free charges are positive or negative or both

The attraction between a charged and an uncharged conductor can be explained by induction provided it is assumed that the forces

between charges diminish with increasing distance apart—an assumption to be tested in chapter 2. Although good insulators are also attracted by a charged body, the experiment of Fig. 1.3 repeated with a split insulator does not show resultant charges on the two pieces afterwards. The inference to be drawn is that insulators contain charges which can move small distances so that attraction still occurs, but that they are bound in equal and opposite amounts so that no splitting of the body can separate two kinds of charge (Fig. 1.4).

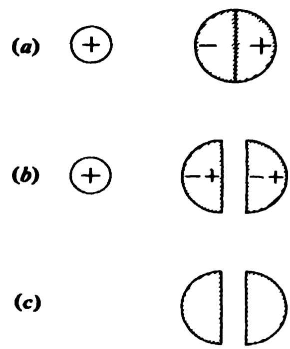


***Fig. 1.4.** Induced charges in insulators: the same operations as in Fig. 1.3 leave two uncharged insulators.*

Production of Charge by Batteries. The disadvantages of the production of charge by friction are that the dampness of the apparatus usually allows charge to leak away (water being a conductor) and that it is difficult to produce a charge of some preassigned magnitude: in short, that steady charges of a required size are not easily available for quantitative experiments. Mechanically-driven machines which will replenish charge as it leaks away can be constructed (section 5.9) but these are inconvenient for our present purpose.

It is found, however, that a battery of voltaic cells connected between the cap and case of an electroscope produces a charge proportional to the number of cells. This charge, moreover, is steady as long as the battery is connected and remains if the connection is broken, although it will now generally leak slowly away. The terminal of the battery giving positive electricity on the cap when connected to it is known as the positive terminal and the other as the negative (Fig. 1.5).

We are not at the moment considering the mechanism whereby charges are separated either by a battery or by friction: our immediate concern has been to find a satisfactory way of producing and measuring charge, and this we have achieved.

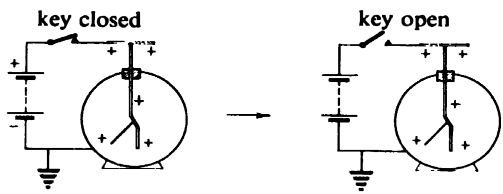


Fig. 1.5. A battery of voltaic cells charges an electroscope (the cap need not now be a container).

1.2 Electric Current

A piece of metallic wire, connected between the terminals of a battery of voltaic cells, rises in temperature; two such wires arranged approximately parallel to each other as in Fig. 1.6 are found to attract or repel according to the way the connections are made to the battery terminals. We recognize by these phenomena, known

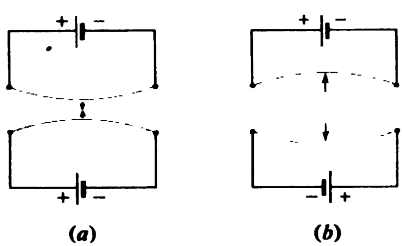


Fig. 1.6. Magnetic forces: like currents attract, unlike currents repel.

as the heating and magnetic effects respectively, the presence of an *electric current* in the wire. It is important to realize that we have no evidence at this stage of a *flow* of anything, and that the term 'current' is merely a name for that in the wire which produces the above effects.

We also find that, if the wire is broken and the ends dipped into the solution of a salt in water, chemical effects occur at the ends of the wires: substances are deposited from the solution (or evolved if

gaseous) or the wires themselves dissolve. This is *electrolysis*, a chemical effect, and it is assumed that the reader is familiar with its terminology and with simple examples of it. The terms *electrode*, *anode* (connected to the positive terminal) and *cathode* (connected to the negative), although originating in connection with electrolysis, are now used in cases where substances other than salt solutions separate the wires.

Any of the three effects can be used to detect a current and in each instance there is a quantity which can be used as a measure of the strength of the current. The heating effect can be used by measuring the heat generated per unit time in a given piece of wire; the chemical effect by measuring the mass of a certain substance deposited per unit time from a certain salt solution; and the magnetic effect by measuring the force on a standard current, although the setting up of this standard current will clearly cause difficulty.

Whichever of the three methods is used, it becomes possible to recognize the existence of a *steady* current—one which does not decay in time—and to establish the important quantitative result that the strength of a steady current in a given arrangement is the same at all points along the wire: for instance, that wherever the wire is broken and a voltameter inserted, the same deposition of mass per unit time occurs. This result is still obtained if several pieces of wire of different materials and cross-sections are connected end to end across the terminals of the battery.

Before we go on to compare the three methods we must look a little more closely at the magnetic effect. If a steady current were easy to maintain, then an arrangement as in Fig. 1.7a would be sufficient as a current-measuring device. The mutual force of attraction between coils A and B in standard positions carrying the two currents is balanced by weights and, if unit current in A is defined as that which attracts B with a force W, then a current in A exerting nW on the *same* current in B is of n units. This arrangement is one form of *current balance*.

Such a steady current for B is, however, impossible to produce over indefinite periods of time and to avoid this difficulty the arrangement is modified slightly. In Fig. 1.7a, if the current in A were first increased until the weight was $2W$ we should say that the current in A had doubled: if then the current in B were increased until the weight was $4W$, keeping the current in A constant, we should say that the current in B had doubled. Thus when both

currents are doubled, the weight increases fourfold: more generally, an m-fold increase in one current and an n-fold increase in the other is accompanied by an mn-fold increase in the force.

If, then, the current to be measured is passed through the two coils in series (Fig. 1.7b) and a balancing weight W represents unit current, an n^2-fold increase in this weight means an n-fold increase

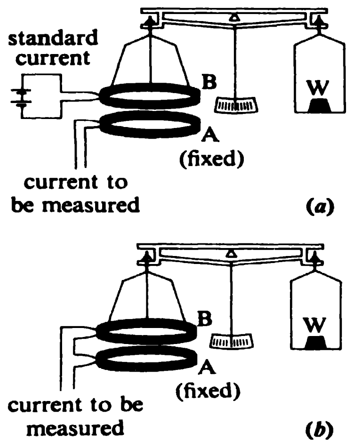


Fig. 1.7. Current balances.

in current. The current, in other words, is proportional to the square root of the force. It cannot be too strongly emphasized that this is true only because we are using the magnetic force as a measure of current—the result is not a law of nature but a consequence of our method of measurement.

Comparison of the various methods of measuring current will be considered from the practical point of view in section 16.1, but we must investigate now whether the three measures of current afforded by the three effects agree. It is found *experimentally* that current measured electrolytically and magnetically are proportional, but that both are proportional to the *square root* of the current measured using the heating effect in metals.

It therefore follows that either we choose

1. Current measured by the magnetic or chemical effect, denoted by I: the heating effect is then proportional to I^2, or
2. Current measured by the heating effect, denoted by I_H: then the magnetic and chemical effects are proportional to $I_H{}^{1/2}$.

It is universally agreed that the former choice be made and that the magnetic effect in particular be used to define a measure for current. Provided this is recognized, we can take as experimental laws the following:

Faraday's law of electrolysis:

$$\text{mass deposited per unit time} = k'I \tag{1.1}$$

Joule's law:

$$\text{rate of generation of heat in a metallic wire} = k''I^2 \tag{1.2}$$

1.3 The Relation between Electric Charge and Current

As we have seen, friction or electrostatic induction will separate equal and opposite charges on two conductors. If these are now connected by a conducting wire the charges disappear and while they are doing so a transient current, detected by its magnetic effect, flows in the wire (Fig. 1.8). Since we know from the

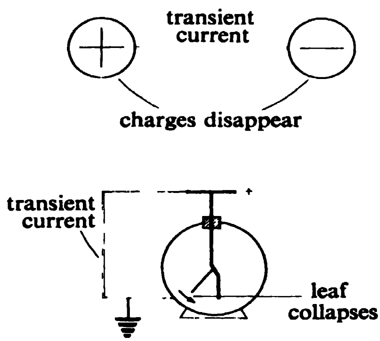


Fig. 1.8. Discharges accompanied by transient currents.

phenomenon of induction that charges in a neutral conductor separate by moving through it, we now conclude that the charges recombine by moving through the connecting wire and that *this movement constitutes the current.*

This picture of the process is the justification for using the word 'current' and we can extend it to cover the case of the steady as opposed to the transient current by realizing that a battery of cells or any other source of such a steady current will constantly replenish the charges at the terminals and keep them constant (Fig. 1.9).

Several questions immediately arise. If the charges at any instant of time t on the conductors in Fig. 1.8 are denoted by $+Q$ and $-Q$ in arbitrary units, the rate of discharge is $\mathrm{d}Q/\mathrm{d}t$. We should like to know how this is related to (*a*) the passage of charge

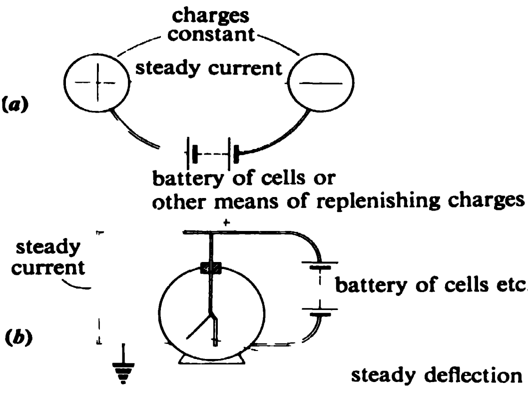


Fig. 1.9. Steady currents.

along the wire and (*b*) the current in the wire as measured by its magnetic effect.

As to the passage of charge along the wire, a reduction of $+Q$ by δQ is accompanied by a reduction of $-Q$ by $-\delta Q$ (Fig. 1.10) since charge is conserved. This could equally well be the result of $+\delta Q$

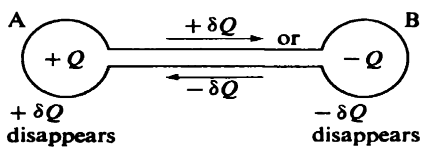


Fig. 1.10. The reduction in size by δQ of the charges on A *and* B *may be due to $+\delta Q$ moving from* A *to* B, *to $-\delta Q$ moving from* B *to* A, *or to the movement of both + and − charges.*

passing from A to B, of $-\delta Q$ passing from B to A, or of the passage of both positive and negative charges. All we can say is that, if flow from left to right is counted positive and from right to left negative, the algebraic sum of the charges crossing any section of the wire must be $+\delta Q$. Thus, if $\mathrm{d}Q/\mathrm{d}t$ is the rate of discharge of A

at any instant, it is also algebraically the rate at which charge crosses any section of the wire

Rowland's Experiment We now turn to the relation between the rate of passage of charge along a wire (dQ/dt) and the current in the wire (I) measured by its magnetic effect (Note that the mere association of I and dQ/dt does not imply their proportionality)

Although the idea of current as essentially something moving rather than as a peculiar condition of matter seems convincing, we have so far no *direct* evidence of any motion accompanying a current Hence, even a qualitative experiment showing that a charged body in motion produces a magnetic effect would be valuable in reinforcing this picture Rowland's experiment, named after its originator, shows the equivalence of currents and moving charges both qualitatively and quantitatively

The apparatus shown in Fig 1 11 was that used by Eichenwald An annular strip of tinfoil with a thin sector removed was fixed on to an insulating disc which could rotate at a known frequency n about an axis through its centre The ring of foil could either be charged and rotated (Fig 1 11b) or be made to carry a current I

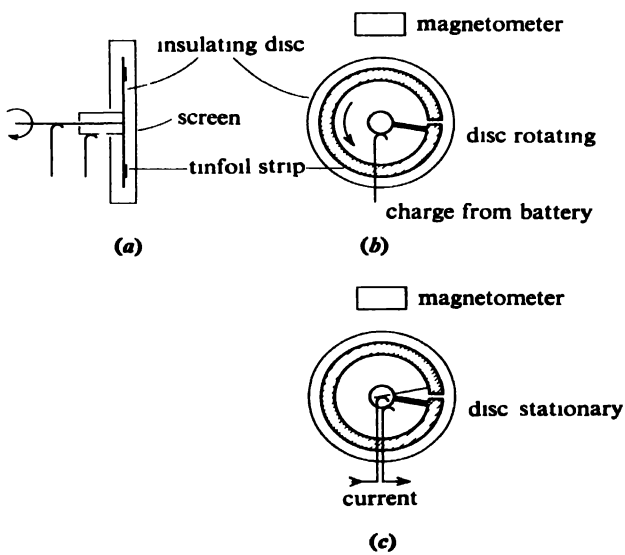


Fig. 1.11. Eichenwald's version of Rowland's experiment. (a) Method of making connections, (b) connections for moving charge; (c) connections for current.

while stationary (Fig. 1.11c). In both instances the magnetic effect was measured by a magnetometer, a current-measuring device equivalent to that illustrated in Fig. 1.7a. The charge Q was measured by an electrometer calibrated in terms of voltaic cells, a method equivalent to that in section 1.1. Figure 1.11a shows the method of making the connections and of screening the disc from external effects (section 4.1).

It was found that the magnetic effect produced by a positive charge rotated in a clockwise sense was the same as that produced by a negative charge of the same magnitude rotated counterclockwise at the same speed, showing that we cannot yet distinguish between these cases. Eichenwald's other results may be summarized as follows: a charge moving with speeds up to 150 m/s produces magnetic effects which in every respect (direction, magnitude, distribution in space) are equivalent to those of an electric current. Table 1.1 gives some of his numerical results recalculated in terms of the description given here, and shows further that the charge passing per unit time, nQ, is proportional to the current, I, whether n or Q is varied.

Table 1.1

EICHENWALD'S VERSION OF ROWLAND'S EXPERIMENT

Q	n	θ	I	$nQ/I \times 10^{-4}$
1,875	77	7·6	6·9	2·11
3,750	77	14·8	13·3	2·17
5,000	77	19·6	17·6	2·18
6,250	77	24·6	22·2	2·16
6,250	15	4·6	4·1	2·27
6,250	30	9·5	8·6	2·19
6,250	100	32·5	29·3	2·13
6,250	150	47·1	42·5	2·21

Q = charge on disc in arbitrary units; n = revolutions of the disc per second; θ = deflection produced on magnetometer by nQ and by I; I = current through foil in arbitrary units.

To summarize, if I is the current in any conductor measured by its magnetic effect and if $\mathrm{d}Q$ is the charge crossing a section of the conductor in an element of time $\mathrm{d}t$ and giving rise to I, then

$$I = k\,\mathrm{d}Q/\mathrm{d}t \tag{1.3}$$

where k is a constant depending only on the particular units chosen.

For direct evidence of association of mass with the moving charge see section 6.9.

1.4 Units of Current and Charge

As we have seen, any arbitrary units may be used to establish laws and as long as everybody agrees to use the same phenomena to measure the quantities involved, the same laws will be discovered. Thus, if all laboratories use a current balance to measure current they will all find that Faraday's and Joule's laws are obeyed: using current balances which differ in their dimensions causes the laws to be expressed with different values for the constants k' and k'' in (1.1) and (1.2) or for k in (1.3).

It is not, however, only the laws themselves which are of interest but the values of the constants. For instance, k' is found to have values characteristic of the substance deposited in electrolysis and is therefore a *property of the substance*; the constant k'' in Joule's law is found to take values characteristic not only of the material through which the current passes but also of the dimensions and the temperature and is thus a *property of the system* at a given temperature; while the constant k in (1.3) does not appear to take different values and is a *universal constant*. Unless we all agree about the units to use we shall not be able to use other people's values of these properties or of the universal constants. The need thus arises for universally agreed units for every measurable quantity.

Our problem concerns current and charge. If we stick to completely independent units for these, then equation (1.3) would stand and k would have a value to be determined by experiment. There is, however, almost universal agreement* to choose units for I and Q so that $k = 1$ and

$$I = \mathrm{d}Q/\mathrm{d}t \tag{1.4}$$

so that we need only agree on a unit *either* of I *or* of Q.

The Ampere, A. The following is the definition of the unit of current now established by international agreement and based on the magnetic effect:

> The ampere is that constant current which, when maintained in two parallel infinitely long rectilinear conductors of negligible circular section placed at a distance of 1 metre apart *in vacuo*, produces between these conductors a force of 2×10^{-7} newton per metre length.

* For an exception, see end of section 16.7 (Coulson).

This definition fixes the value of a constant in the general law of force between currents (section 8.6), just as the choice made before equation (1.4) fixed the value of k at unity. This general law enables a measurement made with one current balance to be related to that made with another and, although we have not yet developed the law, we could, if we chose, go straight on to chapters 7 and 8 to do so. This is effectively what standardizing laboratories such as the National Physical Laboratory do in setting up their current balances to provide us with instruments calibrated in amperes. We may therefore proceed as if we had a current-measuring instrument so calibrated.

The Coulomb, C. Since equation (1.4) is to be used to define a unit of charge from that of current, the coulomb is defined as the charge crossing the section of a conductor in which a current of 1 A flows for 1 s. Measurement of a charge in coulombs, while still possible with a calibrated electroscope, is more often carried out by methods to be considered in sections 5.8 and 16.6.

CGS and MKS Systems. It has long been conventional to use CGS mechanical units throughout physics, but it is more convenient in electricity to use the system based on the metre, the kilogramme and the second (MKS units). If the ampere is chosen as the unit for current, the resultant electric system is known as the MKSA system and is adopted throughout this book. A brief note on the mechanical units is given in appendix 1.2 for those unfamiliar with them. As much help as possible is given to readers who wish to use CGS units, while in section 16.7 a full treatment of the various systems of units in use is given.

1.5 The Nature of Charge

Is electric charge a continuous fluid or are there 'atoms' of electricity as there are of matter? Is charge weightless or is it always associated with mass? Are there distinct positive and negative charges or is one merely matter with a deficiency of the other? These questions relate to the nature of charge and cannot be answered as a result of any phenomena we have so far considered: either alternative is possible in every case. For instance, while the measurements of charge using an electroscope seem to indicate that any magnitude is possible, there could exist elementary charges so small that the addition of one of them produces no detectable change in deflection. Most of this book is concerned with situations in which any such charges produce no noticeable effects, but

on occasion we shall wish to deal with very small masses and very small charges so that answers to the problems posed above must be sought.

It was realized early in the nineteenth century that the laws of electrolysis suggest very strongly what the answers are likely to be. Faraday found that the mass of any substance deposited at an electrode was not only proportional to charge (or current × time) but to the chemical equivalent (atomic weight of the substance divided by its valency) as well. Thus for a substance of atomic weight A relative to hydrogen* and of valency n, the mass deposited by a charge Q would be

$$M = KQAm_{\mathrm{H}}/n \tag{1.5}$$

where m_{H} is the mass of the hydrogen atom and K is a universal constant. Since Am_{H} is the mass of one atom of the substance, M/Am_{H} is the number of atoms deposited and (1.5) becomes

$$\text{Number of atoms of any substance deposited} = KQ/n \tag{1.6}$$

Hence the passage of a given charge Q deposits the same number of atoms of any univalent substance, half the number of any divalent atoms and so on. The inference is that all univalent atoms are associated with the same elementary charge, and n-valent atoms with n times this amount. Moreover, this elementary charge will, from (1.6), have the magnitude $1/K$: we shall denote this (including any sign) by e. However, (1.5) shows that electrolytic deposition will not give K explicitly, but only Km_{H} or its reciprocal e/m_{H}. Experimentally this is found to be $9{\cdot}652 \times 10^7$ C/kg.

Electrolysis thus suggests that electric charge is discrete, that it is associated, at any rate in conduction through electrolytes, with atoms to form what are called *ions*, and that, since deposition may occur on both electrodes, both positive and negative charges are so associated.

Experiments on conduction in gases confirm the existence of ions carrying small multiples of an elementary charge, and also show that a negative particle, known as the *electron*, with e/m 1,836 times larger than e/m_{H}, is a universal constituent of matter. Subsequently, the elementary or *electronic charge* e was determined independently and found to be approximately $1{\cdot}60 \times 10^{-19}$ C for

* Atomic weights are now related to the atomic mass unit (amu) defined as $\frac{1}{12}$ of the mass of the ^{12}C isotope, but this does not affect the argument: m_{H} simply becomes the mass of 1 amu.

both ions and electrons. The electronic mass (about $9{\cdot}11 \times 10^{-31}$ kg) is thus 1,836 times smaller than that of the lightest ion: section 6.9 presents evidence that the charges responsible for conduction in metals are electrons and this would account for the absence of a chemical effect.

This last paragraph is premature in that it deals with material which can only be treated fully later (chapter 11 for determination of e/m, section 3.9 for determination of e). It is included here to give some indication of the sizes of the elementary charge and the associated masses so that we can assess whether it is reasonable to treat some distributions of charge as continuous and virtually massless. It also makes clear that the law of conservation of charge is dependent on (*a*) a constant electronic charge and (*b*) a constant difference between the numbers of positive and negative elementary charges in any isolated system. Both (*a*) and (*b*) are verified by atomic and nuclear experiments.

1.6 Charge Densities and the Point Charge

The view we take of a charge spread over a finite volume depends on our distance from it. If the distance is large compared with a linear dimension of the volume, the total charge Q is effectively concentrated at a point and in the limit we should arrive at the concept of a *point charge*. The importance of this idea lies in the simplicity we can expect in the laws relating to it, since complex factors like the shape and size of the charge distribution have been eliminated.

When our distance is of the same order of magnitude as the size of the volume, and if both are large compared with interatomic distances, the charge distribution will behave as if it were continuous. If a volume $\delta\tau$ contains a charge δQ, the mean density of charge $\bar{\rho}$ within the volume is given by

$$\delta Q = \bar{\rho}\,\delta\tau \qquad \text{(Definition of } \bar{\rho}\text{)} \quad (1.7)$$

If $\delta\tau$ be shrunk to an infinitesimal volume $d\tau$ around a point, then in this limit the charge density at the point, ρ, is defined by

$$dQ = \rho\, d\tau \qquad \text{(Definition of } \rho\text{)} \quad (1.8)$$

The total effect of a charge Q can be obtained by integrating the effects of the dQ's (the principle of superposition: section 2.1).

If the charge is spread thinly over a surface, a more suitable quantity would be the surface density of charge, σ, defined by

$$dQ = \sigma\, dS \qquad \text{(Definition of } \sigma\text{)} \quad (1.9)$$

where dQ is the charge occupying an area dS. This is still a volume distribution and if the thickness is b (Fig. 1.12) it is easy to show that the relation between the densities is $\rho b = \sigma$, but in cases where b is much smaller than any other linear dimension σ is the more convenient.

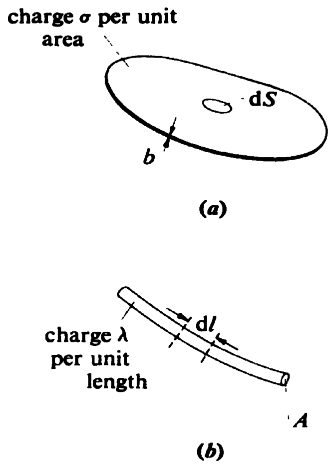


Fig. 1.12. (a) Surface density of charge; (b) line density of charge.

In a similar way, a line charge density λ defined by

$$dQ = \lambda \, dl \qquad \text{(Definition of } \lambda) \qquad (1.10)$$

is often more convenient where one linear dimension is much larger than any other. Again, it is easy to show that $\rho A = \lambda$ for a line density with an area of cross-section A.

When we come to distances from a charge distribution comparable with the distances between the elementary charges themselves, then no longer can we treat it as a continuum with smooth densities of charge and such cases require great care (e.g. sections 13.8 and 13.9).

Finally, it may be necessary to express a charge or charge density in terms of the elementary charges of which they are composed. If the elementary charge is denoted by e, then

$$\rho = ne \qquad (1.11)$$

where n is the number of charged particles per unit volume. If there are several different types of particle with densities $n_1, n_2, \ldots$ and charges $e_1, e_2, \ldots$, some of which may be negative, then

$$\rho = \sum n_i e_i \qquad (1.12)$$

1.7 Current Densities and the Velocity of Charges

We are now satisfied that any moving charges constitute an electric current and that, since charge is known to be discrete, its flow will be discontinuous. Nevertheless, when there are great numbers of very small charges with sufficient velocities the current can be treated as if it were the flow of a fluid.

So far we have concentrated on currents flowing in wires, but it is clear that they may flow in conductors of large volume and we can also envisage currents composed of the elementary charges themselves passing across an evacuated space and not bounded by any conducting surface. In all these more general examples we must remember that the total current is given by the rate of passage of charge *across a specified area* even though the area may be understood, as in the case of a wire, rather than explicitly stated.

Figure 1.13 shows that across an area like $A_{\parallel}$ the current is zero, while across A_n it is a maximum for a given flow of charge. The *current density* at a point is defined by taking the current per unit area *normal* to the flow. Thus, in Fig. 1.13, if the area A_n is shrunk

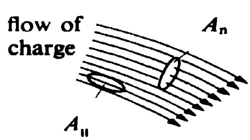


Fig. 1.13. Current is associated with an area. Current across $A_{\parallel} = 0$ since no charge crosses it.

to dA_n about a point until the current across it is dI, the current density J is given by

$$dI = J\, dA_n \qquad \text{(Definition of } J) \qquad (1.13)$$

We shall see later (appendix 4.1) that J is a vector quantity while I is not.

If a current flows in a thin surface layer as in Fig. 1.14, the *surface*

current density J_s is a more useful quantity. It is defined by

$$dI = J_s \, dl \qquad \text{(Definition of } J_s\text{)} \qquad (1.14)$$

and is thus the current per unit length.

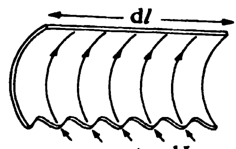


Fig. 1.14. Surface current density is the current per unit length.

Velocities of Charge. It is sometimes useful to be able to express currents and current densities in terms of the velocities of the elementary charges. If there are n per unit volume each with a charge e, then Fig. 1.15 shows that the charge crossing an area A in unit time is that contained in a cylinder of base area A and vertical

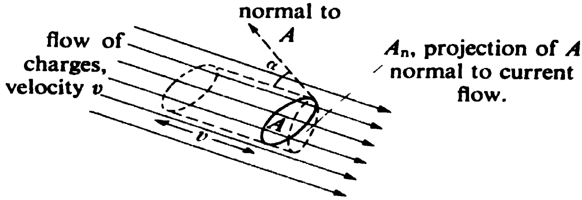


Fig. 1.15. Velocity of charges and its relation to current.

height $v \cos \alpha$, v being the velocity. The current I is therefore $nevA \cos \alpha$ or

$$I = nevA_n = \rho v A_n \qquad (1.15)$$

where A_n is the area projected normal to the current flow. By equation (1.13),

$$J = nev = \rho v \qquad (1.16)$$

More generally, a current will consist of several types of charged particle with densities $n_1, n_2, \ldots$, charges $e_1, e_2, \ldots$ and velocities $v_1, v_2, \ldots$. Equations (1.15) and (1.16) then become

$$I = A_n \sum nev = A_n \sum \rho v \qquad (1.17)$$

and

$$J = \sum nev = \sum \rho v \qquad (1.18)$$

In these relations both e and v (or ρ and v) may be either positive or negative, and the summations are algebraic. Negative charges with negative velocities thus give positive currents, in line with the results of Eichenwald's experiment.

More generally still, it is possible that each set of charged particles has a range of velocities including (*a*) random velocities whose mean is zero and (*b*) *drift* velocities whose mean is the v of (1.15)–(1.18). Currents encountered in normal laboratory experiments have charges whose drift velocities are of the order of 0·01 cm/s only (problem 1.13).

1.8 Summary of Chapter 1

If this chapter has appeared straightforward we should realize that the work of centuries has been compressed into a few pages (see appendix 1.1). The concepts of charge and current presented here were only developed by the work of many outstanding men and some of the possible lines of investigation which they had to eliminate have been omitted. For instance, the laws of electrolysis have no mention of the concentration of the electrolyte solution or of the size of the electrodes: it is not obvious that these would have no effect on the mass deposited and Faraday had to eliminate them by experiment. But to include every detail of this sort would be both wearisome and a disservice to the work of people like Faraday which was done so that we could avoid their sidetracks. We are concerned, moreover, with the subject as it stands today and it is right that our account should be more direct than a strictly historical one.

In this chapter we have established as many of the properties of electric charge and current as are needed to recognize and measure them. The measurement has enabled the following quantitative laws to be established:

1. The conservation of charge
2. $I = \mathrm{d}Q/\mathrm{d}t$

The constant value of a steady current at all points along a wire, first established by experiment, now appears as a deduction from these two. Joule's law (only obeyed by a limited range of materials) and Faraday's law of electrolysis are of less general importance than 1 and 2.

Charge and current densities have been defined and the relations $I = nevA_n$, (1.15) and $J = \rho v$, (1.16), deduced from law 2 above.

We have also considered enough evidence to show that the dis-

creteness of charge on an atomic scale is likely, although for many problems charge may still be regarded as continuous just as matter itself often is.

We cannot proceed usefully any further until we have looked at the quantitative laws governing the electric force between charges and the magnetic force between currents. Which is taken first is a matter of taste, but we expect the interaction between stationary charges to be less complex than one between moving charges, where velocities are an additional variable. We shall therefore consider first (chapters 2–5) the electric force and then, before tackling the magnetic force, we shall look at the production and properties of steady currents (chapter 6) using many of the ideas by then developed.

Appendix 1.1 Historical Note

Although the ancient Greeks were aware that amber (ἤλεκτρον) could be electrified by rubbing, it was not until 1600 A.D. that William Gilbert (1540–1603) showed that many other substances behaved in the same way, and only in 1731 did Stephen Gray (d. 1736) announce the discovery of conductors of electricity. Charles-François du Fay (1698–1739) distinguished vitreous and resinous electricity in 1733, while 14 years later Benjamin Franklin (1706–1790) suggested the law of conservation of charge and the use of + and − to distinguish two different charged conditions. From Gilbert onwards there was much speculation about the nature of electricity, but little quantitative work of any description was carried out until late in the eighteenth century: this will be discussed in the next chapters.

Luigi Galvani (1737–1798) announced in 1791 the discovery of what was called animal electricity—the convulsion of a frog's leg when a connection was made across its ends through two dissimilar metals. But Alessandro Volta (1745–1827) showed with his pile (of copper, zinc and moistened pasteboard discs repeated in that order) that the animal was not essential and that the pile produced effects similar to those of frictional electricity, viz., the charging of an electroscope and the production of electric shocks and sparks. These last two phenomena associated with discharge were then much used for identifying electricity even though they do not lend themselves to exact measurement.

Current from the voltaic pile enabled the chemical effect to be studied closely, first by Nicholson and Carlisle in 1820 and most

fruitfully by Michael Faraday (1791–1867) who stated his laws of electrolysis and conjectured the atomicity of electricity in 1833. The magnetic effect as we have studied it was discovered by André-Marie Ampère (1775–1836) in 1820, while the heating effect was first measured by James Prescott Joule (1818–1889) in 1841. The first experiment to determine whether moving electrified bodies were equivalent to currents was carried out by H. A. Rowland (1848–1901) in 1876 and was subsequently repeated several times by him and others between then and 1903.

The discovery of the electron was largely the result of the work of J. J. Thomson (1856–1940) who measured the specific charge (e/m) for a number of charged particles and in the period 1897–1899 established the basis of atomic physics summarized in section 1.5.

Appendix 1.2 The MKS System of Units

The system of units most commonly used by physicists has been the CGS system which includes the dyne as the unit of force and the erg as the unit of work and energy. This system can easily be extended to electricity and magnetism, but complications arise both because more than one consistent CGS system is used and because in experimental work 'practical' units (those in which instruments are calibrated) are employed and these are rarely the same as the CGS units.

If, however, a system of units based on the metre, kilogramme and second is used instead, we find that nearly all the practical electric and magnetic units can be included, thus avoiding the

Table 1.2

MKS UNITS IN MECHANICS

Quantity	*CGS unit and symbol*	*MKS unit and symbol*	*Relation*
Mass	gramme, g	kilogramme, kg	1 kg $= 10^3$ g
Length	centimetre, cm	metre, m	1 m $= 10^2$ cm
Time	second, s	second, s	
Velocity	cm/s	m/s	1 m/s $= 10^2$ cm/s
Linear momentum	g-cm/s	kg-m/s	1 kg-m/s $= 10^5$ g-cm/s
Density	g/cm^3	kg/m^3	1 kg/m$^3 = 10^{-3}$ g/cm^3
Force	dyne, dyn	newton, N	1 N $= 10^5$ dyn
Work, energy	erg	joule, J	1 J $= 10^7$ erg
Power	erg/s	watt, W	1 W $= 10^7$ erg/s
Moment of force, couple	dyn-cm	N-m	1 N-m $= 10^7$ dyn-cm
Angular momentum	g-cm^2/s	kg-m^2/s	1 kg-m^2/s $= 10^7$ g-cm^2/s
Pressure	dyn/cm^2	N/m^2	1 N/m$^2 = 10$ dyn/cm^2

necessity for different systems in theoretical calculations and experimental measurements.

Since many readers will be unfamiliar with MKS units in mechanics, table 1.2 presents a summary of them and their relation to CGS units. The only unfamiliar name is likely to be the newton —the force which gives to a mass of 1 kg an acceleration of 1 m/s^2: it is 10^5 dynes or about 100 g wt. The unit of work would be the newton-metre, equal to $10^5 \times 10^2$ or 10^7 erg, already familiar as the joule.

References

For more detailed descriptions of very elementary electricity and magnetism a book such as that of Mitchell (1960) or one of equivalent standard should be consulted. The historical bare bones in appendix 1.1 can be covered by selective reading of Whittaker (1951): more directly, Magie (1964) gives extracts from original papers of many of the pioneers, while the *Experimental Researches of Faraday* (1951) gives a fascinating account of investigations relevant to this chapter.

PROBLEMS

SECTION 1.1

1.1 Devise a method for measuring charge based on the force between charges and similar to that used for measuring current in the current balance of section 1.2. Do you need to assume conservation?

1.2 Revise the explanations, in terms of the movements of charges, of the charging of an electroscope both by contact and by induction.

1.3 You are shown two identical rods which repel each other when placed end to end, but which both show attraction for resinous and vitreous electricity. You are *told* that the rods are of Spurite and that they are charged with a third form, spurious, of electricity. What do you suspect to be the truth, and how would you test your suspicion?

SECTION 1.2

*1.4 Assess the relative merits of the three possible methods of measuring current. Consider both the practical and theoretical aspects.

SECTION 1.3

*1.5 Some authors *define* current as rate of passage of charge so that $I = \mathrm{d}Q/\mathrm{d}t$ is a definition of I. How would this approach affect the treatment in sections 1.1, 1.2, 1.3: in particular, is Rowland's experiment superfluous?

SECTION 1.4 AND APPENDIX 1.2

1.6 (A simple problem in MKS units) What force is needed to impart to a mass of 8 kg an acceleration of 2·5 m/s^2? If the force is applied constantly over a distance of 5 m, what is the increase in the kinetic energy of the mass? If the mass starts from rest, what are its final velocity and momentum? Also work out the problem in CGS units.

1.7 Find the values of the following constants in MKS units: the density of water at 4°C; the specific heat of water at 15°C; the surface tension of water at 20°C; the density of air at S.T.P.; the acceleration due to gravity.

SECTION 1.5

1.8 If it were possible to charge a conductor weighing 10 g with 1 C, what change in mass would occur on discharge to earth? How does this change compare with that detectable by a chemical balance? (As to the feasibility of a charge of 1 C, see problems 2.3 and 4.5.)

1.9 A gramme-atom of an element is a mass of it containing the same number of grammes as its atomic weight relative to hydrogen. The gramme-molecule (or mole), the gramme-ion and gramme-equivalent are similar. Show that all gramme-atoms of any element contain the same number, N_A, of atoms, and that all gramme-molecules contain N_A molecules etc. (N_A is Avogadro's number.) Find the value of N_A from the data in section 1.5. Show that a substance of density ρ and atomic weight M has $N_A\rho/M$ atoms per unit volume.

1.10 The Faraday, F, is that quantity of charge which deposits 1 gramme-equivalent of a substance in electrolysis. Show that F is a universal constant equal to N_Ae and find its value from the data in section 1.5.

1.11 If formulae are intended for use with MKS units, the symbol N_A must stand for the number of atoms per kg-atom, etc., and F for the quantity of charge depositing 1 kg-equivalent. What are then the numerical values of N_A and F?

SECTION 1.6

1.12 Show, with the symbols used in section 1.6, that the relations between σ and λ and the equivalent ρ are $\rho b = \sigma$ and $\rho A = \lambda$.

SECTION 1.7

1.13 Assuming that each copper atom contributes one free electronic charge to the current in a wire, *estimate* the mean drift velocity of charges when the wire has a diameter of 1 mm and carries a current of 1 A. (Atomic weight of copper = 63·6; density of copper = 8·9 g/cm^3.)

CHAPTER 2

THE LAW OF FORCE BETWEEN CHARGES (*IN VACUO*)

We now go on to investigate more fully the force between charged bodies described in section 1.1. Such a force should depend, among other factors, on the size, shape and orientation of the bodies if their sizes are of the same order of magnitude as their distances apart. If we look instead for a law of force between what are effectively two *point* charges, the only factors involved ought to be the magnitude and sign of the charges, their distance apart and the direction of the line joining them, and the law should be correspondingly simple. Because few problems concern only *two* point charges we shall want to know how these electric forces combine when, for instance, both of two point charges exert forces on a third: once this is established, the law can be used to find forces between large bodies carrying charge by the usual process of considering elements of volume, and integrating.

To establish the complete law, we shall examine first the direction of the force and the combination of forces; secondly, the dependence on magnitude and sign of the charges; and thirdly, the dependence on distance. The law will then be applied to particular problems, with an eye on any limits in its range of application.

2.1 Direction of the Force and Superposition

Two point charges situated at O and P, as in Fig. 2.1, will exert forces on each other which must, by Newton's third law, be equal and opposite. Since there is, in addition, symmetry about the line OP, the force F_{OP} (acting on P due to O) must act along the joining

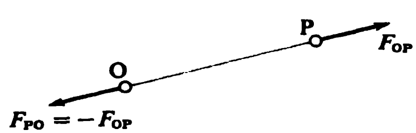


Fig. 2.1. Central forces. F_{OP} passes through O wherever P is situated.

line and so must F_{PO}. The force system will therefore be as indicated in Fig. 2.1 if the charges are of the same sign. The result means that the electric force falls into the very important class of *central forces*: that is, forces whose lines of action always pass through their point of origin. Thus, wherever P was situated, the force on it due to O would always pass through O.

Crude experimental verification can be obtained by suspending from insulating threads two light conductors, small compared with their distance apart, and charging them. The mutual force causes a displacement of each which is, as far as can be detected, in the same direction in space as the line joining the centres, whatever the sign of the charges.

This part of the general law is as important as the others to be considered immediately but is often neglected, possibly because it appears self-evident. It should be emphasized, however, that the symmetry about OP which allows us to deduce the result assumes that charge is a *scalar* quantity (see problem 2.1 and appendix 2.1). The experiments of section 1.1 show that charge can be specified by a single number and thus justify the assumption.

Connected with the direction of the force is the question of superposition. This can best be explained with reference to Fig. 2.2

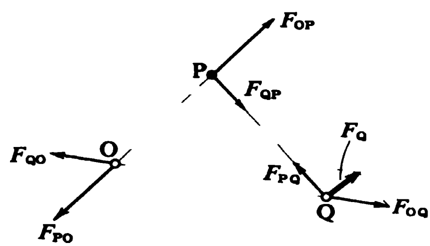


Fig. 2.2. Superposition of electric forces (see problem 2.1b).

in which O, P and Q are three charges, F_{OP} and F_{PO} are the forces between O and P *in the absence of Q*, and similarly for the other forces. If the resultant force on Q when all three charges are present is the vector sum (see appendix 2.1) of F_{PQ} and F_{OQ}, then these forces are said to superpose or to obey the *principle of superposition*. In elementary mechanics we can show that elastic and gravitational forces obey the principle by verifying experimentally the parallelogram law of addition, but this does not mean that electric forces will necessarily behave in the same way. Direct

experiments can be performed by finding the resultant displacement of a small charged body Q, suspended again by a light thread, due to two other charged bodies O and P as in Fig. 2.2. Such experiments are subject to errors due to leakage and the finite size of the bodies, but do suggest that superposition holds.

2.2 Dependence on Magnitude and Sign of Charge

Had we chosen to define a unit of charge in terms of the force exerted on a standard charge at a standard position, much as we did with current, then it would follow that the force between charges of magnitude Q_1 and Q_2 a fixed distance apart would be proportional to Q_1Q_2. We have chosen a more practical method of measuring charge (by means of an electroscope) and must therefore appeal to an experiment like that of Henry Cavendish (1731–1810) who appears to have been the first to show that the force between two equally charged bodies varies directly as the square of the charge measured on an electroscope. His method used the fact that a charge Q on one conductor is shared equally when an identical conductor is made to touch the first: this follows from symmetry and the conservation of charge and is easily checked by an electroscope.

We shall assume that the force is directly proportional to the product of the magnitudes of the charges and shall justify the assumption by testing accurately a consequence of it later (section 5.2). If the sign of the charges is included in the product, the sign of the force conforms to the normal convention—the positive repulsive force F_{OP} in Fig. 2.1, for instance, is in the direction of positive displacement from O. There is no evidence that the sign of a charge affects any other part of the law.

2.3 The Inverse Square Law

It is well known from the theory of gravitation that a mass inside a uniform spherical shell of matter experiences no resultant force and that this can be shown to follow from the inverse square law. Joseph Priestley (1733–1804) observed that charged bodies inside charged hollow conductors apparently experienced no force and he conjectured that the electric force obeyed a similar law. Charles Augustin Coulomb (1736–1806) in 1785 demonstrated the inverse square law directly both for repulsion, with his torsion balance, and for attraction, with an oscillation method (see problem 2.2). The

confidence of theoretical workers in the law was based mainly on his results.

Our confidence in it lies rather in a series of experiments designed to check Priestley's observation, the first being carried out by Cavendish in 1772 but not published until 1879, a second by Clerk Maxwell (1831–1879), and a third by Plimpton and Lawton in 1936.

To see the point of the Cavendish experiment, as it is called, we shall first show that, if from Coulomb's rough measurements it is assumed that the law is an inverse nth power law, a small charge inside a charged conducting spherical shell experiences in general a force towards or away from the centre; except that when $n = 2$ exactly, no force exists.

For this purpose consider a small charge Q inside a sphere, centre O (Fig. 2.3), and divide the sphere into two segments by a plane through Q perpendicular to OQ. The charge on the sphere will,

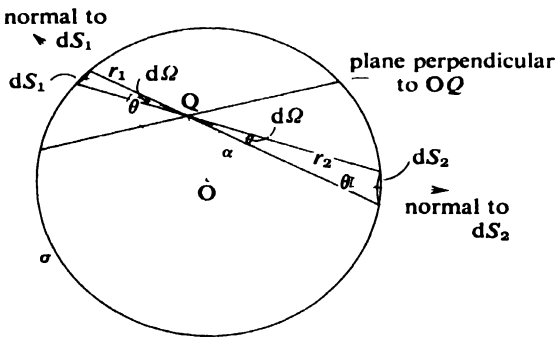


Fig. 2.3. Deduction of the force on Q inside a spherical shell of charge. Areas dS_1 *and* dS_2 *are exaggerated for clarity.*

because of the symmetry, spread uniformly over the surface with a density σ as long as Q is sufficiently small to produce an insignificant change in the distribution. Consider first the force on Q due to two small surface elements dS_1 and dS_2, both cut off by a solid angle $d\Omega$ with apex at Q, and let dS_1 refer to the element on the smaller segment so that $r_1 < r_2$.

The force on Q due to the charge on these elements is

$$dF \propto \sigma\, dS_1\, Q/r_1^{\,n} - \sigma\, dS_2\, Q/r_2^{\,n}$$

which, since $d\Omega=(dS_1 \cos\theta)/r_1^2=(dS_2 \cos\theta)/r_2^2$ (appendix 2.3), is also

$$dF \propto \frac{\sigma Q\, d\Omega}{\cos\theta}\left(\frac{1}{r_1^{n-2}}-\frac{1}{r_2^{n-2}}\right)$$

the direction of this being towards dS_2 if both σ and Q are positive and $n>2$. All pairs of surface elements whose joining line makes the same angle α with OQ produce similar forces directed into the larger segment and the components along the plane perpendicular to OQ will cancel, leaving a resultant force towards the centre O. The whole sphere can be divided in this way and we arrive at the following conclusions:

$n>2$, σ +, the force on a positive Q is towards the centre, on a negative Q away from the centre,
$n<2$, σ +, the force on a positive Q is away from the centre, on a negative Q towards the centre,
$n=2$, no force.

Hence, unless $n=2$, if we place a conductor inside the sphere and connect it to the sphere by a metallic wire, there will be a force on the free charges we know to exist in the wire: these will move to or from the inner conductor which will be left charged when the wire is removed.

In all experiments of this type, any charge left on the inner conductor was less than the smallest detectable by the most sensitive methods available at the time. This lower limit to the detectable charge determines the amount by which n can be said to differ from 2. The calculation of this amount is easy if the inner conductor is spherical and concentric with the outer, and Plimpton and Lawton showed that the calculation was not materially affected even if the lower half of such a conductor were replaced by detecting apparatus. Cavendish was able to show that $n=2\pm 1/50$, and Maxwell that $n=2\pm 1/21{,}600$ using a more sensitive electrometer: this is the best that can be achieved if contact is made and broken because contact potentials (section 12.7) are set up. To avoid these, Plimpton and Lawton mounted the detecting apparatus inside the outer sphere and connected the inner and outer conductors permanently together. The outer was continuously discharged and recharged with alternate signs at 2 c/s by a specially designed generator, the low frequency being necessary to minimize inductive effects (Fig. 2.4). This experiment showed that $n=2\pm 1/10^9$.

To avoid confusion later, it should be emphasized that we have demonstrated the absence of force on a charge inside a hollow *sphere* of charge and have shown that this occurs only if the law is inverse square. We shall later (section 4.1) be able to generalize

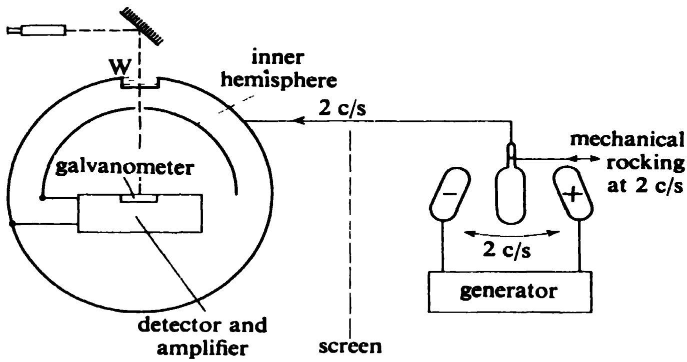


Fig. 2.4. Plimpton and Lawton's apparatus. W is a window of salt solution to ensure that the outer shell is a complete conducting surface.

this—that if the law is inverse square, then there is no force on a charge inside a hollow conductor *of any shape*. We could not use a shape other than a sphere in our proof above because we have as yet no means of finding how the charge distributes itself over a conductor with such a shape.

2.4 Forms of Coulomb's Law

The laws established with varying degrees of reliability in the previous three sections show that the force **F** between two point charges Q_1 and Q_2 separated by a distance r is proportional to Q_1Q_2/r^2 and is along the joining line. This we shall refer to as *Coulomb's law*, summarized by

$$F = kQ_1Q_2/r^2 \quad \text{in magnitude, along } r \text{ in direction (see Fig. 2.5)} \tag{2.1}$$

where k depends only on the units chosen for the various quantities and is thus a universal constant (section 1.4).

It is often convenient to choose a unit for Q which makes $k = 1$ and, if F and r are in dyne and cm respectively, such a unit for Q is called the CGS electrostatic unit (e.s.u.) of charge: other names in use for this unit are the franklin (Fr) and the statcoulomb.

However, we have already chosen for ourselves the coulomb as the unit of charge and the MKS mechanical units: k is thus a quantity to be determined in the MKSA system by experiment. It is usual to write k as $1/4\pi\varepsilon_0$ where ε_0 now becomes the universal

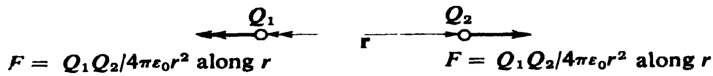


Fig. 2.5. Coulomb's law. For Q_2, the origin is at Q_1 and r is directed from Q_1 to Q_2; and vice versa.

constant to be determined by experiment. The reasons for choosing a more complicated expression are the following: the factor ε_0 is in the denominator rather than the numerator because we then find that a property of materials known as the relative permittivity (chapter 13) always occurs as a multiplier of ε_0; while the factor 4π makes the system we use the so-called *rationalized* MKSA system (its omission would give us a *non-rationalized* system). This is the one recommended for adoption by successive international conferences on weights and measures, and it has certain advantages which we shall review in chapter 16.

Coulomb's law can thus be expressed in the form

$$F = Q_1Q_2/4\pi\varepsilon_0r^2 \quad \text{in magnitude, along } r \text{ in direction} \qquad (2.2)$$

In order to avoid the addition of the words denoting direction while retaining an indication that the force is a vector, the notation outlined in appendix 2.1 is used and the law written

$$\mathbf{F} = Q_1Q_2\hat{\mathbf{r}}/4\pi\varepsilon_0r^2 \qquad (2.3)$$

where $\hat{\mathbf{r}}$ is a unit vector along r, or

$$\mathbf{F} = Q_1Q_2\mathbf{r}/4\pi\varepsilon_0r^3 \qquad (2.4)$$

in which numerator and denominator of (2.3) have been multiplied by r.

The principle of superposition is introduced by writing the force on a charge Q due to a collection of others, $Q_1, Q_2, \ldots$ (Fig. 2.6) as

$$\begin{aligned} \mathbf{F} &= \mathbf{F}_1+\mathbf{F}_2+\mathbf{F}_3+\cdots \\ &= Q(Q_1\mathbf{r}_1/4\pi\varepsilon_0r_1{}^3+Q_2\mathbf{r}_2/4\pi\varepsilon_0r_2{}^3+\cdots) \end{aligned} \qquad (2.5)$$

Value of ε_0. The constant ε_0, variously known as the *permittivity of free space*, the *electric constant* or simply as *epsilon nought*,

takes the value $1/4\pi$ when CGS e.s.u. are used. Its value in MKSA units can be determined in principle by measuring the force between two point charges of known magnitude and distance apart, but this is not a practicable method. Instead, Coulomb's law is used to deduce formulae which lend themselves more readily to measurements and the results of these show that $4\pi\varepsilon_0$ has a value very close to $10^{-9}/9$ (sections 5.8 and 16.7).

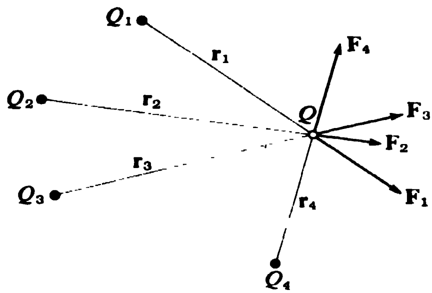


Fig. 2.6. The resultant force on Q due to a collection of charges is the vector sum of the **F**'*s.*

Formulae derived from Coulomb's law can thus be used either with CGS e.s.u. or with MKSA units, provided the following substitutions are made:

$$\left.\begin{array}{lll} 4\pi\varepsilon_0 = 1 & \text{for CGS e.s.u.} \\ 4\pi\varepsilon_0 \approx 10^{-9}/9 & \text{for MKSA units} \end{array}\right\} \qquad (2.6)$$

2.5 Range of Application of Coulomb's Law

A law is usually established experimentally over a limited range of the variables included in it and is then applied with caution outside this range, with an eye open for disagreements with experiment which might indicate its breakdown. Coulomb's law is no exception and we consider in turn several limitations imposed by our method of establishing it.

First, it has not been verified, as far as we are concerned, for huge values of Q or, perhaps more important, for very small ones. In particular, we have by no means verified it for single elementary charges since our conductors must contain very many of them and our result may be the effect of averaging some different law over a vast number.

Secondly, the distances used have been in the centimetre–metre range and any extrapolations to atomic or astronomical distances are tentative. Mason and Weaver (1929), for instance, have pointed out that if the law were such that the force was proportional to $(e^{-a^2/r^2})/r^2$, where a is a length comparable with the size of an atom, then for values of r of a centimetre or so the law would be indistinguishable by experiment from inverse square, but at atomic distances would be very different. A similar argument clearly applies to $e^{(r^2/a^2)}/r^2$ where a is now of astronomical size.

Thirdly, the charges in our experiments were stationary relative to each other: hence the term *electrostatic* to describe the forces. But what happens if the charges move? We know from chapter 1 that the moving charges constitute currents between which a magnetic force operates, but this is in addition to the electrostatic force which must still be acting. What we don't know is whether this electrostatic force is unaltered by the motion. Moreover, the charges in the experiments are only stationary taken as a whole: there may be considerable random motion of the elementary charges present and once again Coulomb's law may be only an average over a period of time of some more fundamental law.

Finally, we have assumed that the presence of air between the charges can be ignored—a point we shall take up when considering the effect of a medium on electric forces (chapter 13).

Lest all this seem too gloomy a prospect for a law we have only just established, it is reassuring to note that it is still applied today with some confidence over wide ranges of the variables, and with success. For the moment, we shall be making two important assumptions: that Coulomb's law is valid for atomic distances and for moving charges. Consequences of these can later be compared with experiment (sections 4.3 and 11.1).

Problem 2.6 illustrates the enormous range of forces existing in nature by a comparison of gravitational and electric interactions. When short-range nuclear forces are included, the range extends even further.

2.6 Applications of Coulomb's Law

The calculation of the force exerted by one set of effectively point charges on another is carried out in a straightforward way by vector addition of the forces (Fig. 2.7), although in particular cases more powerful and neater methods are available (see for example section 4.6).

In the same way, if we are given two continuous distributions of charge, then we can find the force between them by splitting both into small elements of volume: the forces between these elements can then be resolved and integrated over the volume of each distribution. This is no different in principle from the method for

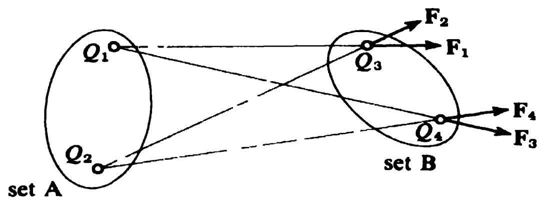


Fig. 2.7. Forces between sets of charges. The resultant force on set B *is the vector sum of* $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ *and* $\mathbf{F}_4$.

discrete charges. In practice, however, we are only likely to know the distribution in advance for conductors with high symmetry like the sphere in section 2.3, but there are also theoretical problems where it may be important to be able to calculate the force between *assumed* distributions (for example, in atomic models—section 4.3).

We shall find problems concerning continuous distributions more easily treated by methods to be developed in the next chapters, but

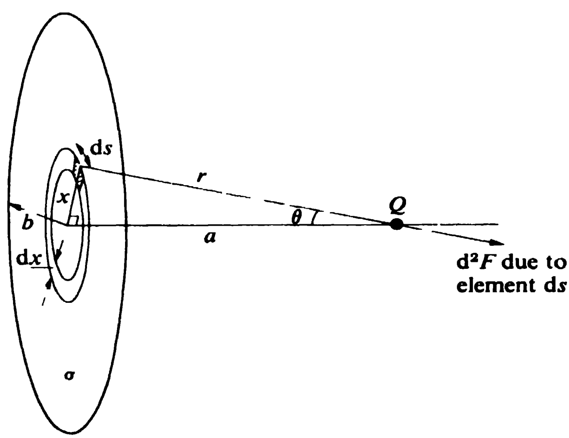


Fig. 2.8. Calculation of the force on Q due to a plane sheet of charge.

as an example of the above outline let us find the force between a charge distributed uniformly over a large plane circular sheet with a surface density σ and a single point charge Q situated a perpendicular distance a from the plane. In Fig. 2.8 we have first divided the plane into annular strips, a typical one having internal and external radii x and $x+\mathrm{d}x$ respectively. All the charge in the annulus lies at a distance r from Q, and so an element of the ring of length ds carrying a charge σ dx ds exerts a force

$$\mathrm{d}^2F = \frac{Q\sigma\,\mathrm{d}x\,\mathrm{d}s}{4\pi\varepsilon_0 r^2} \quad \text{along } r$$

The force due to the various elements of the ring form a cone of semi-vertical angle θ at Q, and their components perpendicular to a will cancel. The component parallel to a is $\mathrm{d}^2F\cos\theta$ for every element of length ds, and hence for a total length of ring $2\pi x$ the force is

$$\begin{aligned}\mathrm{d}F &= \frac{Q\sigma\,2\pi x\,\mathrm{d}x\cos\theta}{4\pi\varepsilon_0 r^2} \quad \text{along } a\\ &= \frac{Q\sigma\,ax\,\mathrm{d}x}{2\varepsilon_0 r^3} \quad \text{along } a\end{aligned}$$

If b is the radius of the plane sheet of charge, the total force on Q is

$$\begin{aligned}F &= \frac{Q\sigma a}{2\varepsilon_0}\int_0^b \frac{x\,\mathrm{d}x}{(a^2+x^2)^{3/2}}\\ &= \frac{Q\sigma a}{4\varepsilon_0}\int_{x=0}^{x=b}\frac{\mathrm{d}(a^2+x^2)}{(a^2+x^2)^{3/2}}\\ &= \frac{Q\sigma a}{2\varepsilon_0}\left[-\frac{1}{(a^2+x^2)^{1/2}}\right]_{x=0}^{x=b}\\ &= \frac{Q\sigma a}{2\varepsilon_0}\left[\frac{1}{a}-\frac{1}{(a^2+b^2)^{1/2}}\right].\end{aligned}$$

For a very large plane, $b\to\infty$ and in the limit

$$F = Q\sigma/2\varepsilon_0 \tag{2.7}$$

independent of a. This is a result we shall be looking at again.

2.7 Mutual Potential Energy of Charges

The potential energy of a system is the energy associated with its position relative to other systems and with the relative positions of its parts: it is defined as the work done by forces *external* to the system in taking it from some standard configuration to the given one. The standard configuration thus forms a zero which must be stated whenever a potential energy is quoted. For instance, with sea-level as a zero the potential energy of a mass m a small height h above sea-level is mgh.

With a pair of point charges the natural zero would seem to be when they are so far apart that their interaction is negligible —an infinite distance. Hence, for the potential energy of $+Q_1$ and $+Q_2$ separated by r, we need to find what work would be done by an external force in bringing say Q_2 from infinity to its final position while Q_1 remains fixed.

We need not worry about the work done by the external force needed to keep Q_1 in position: its point of application is stationary and the work is zero. Referring to Fig. 2.9, the path taken by Q_2 is

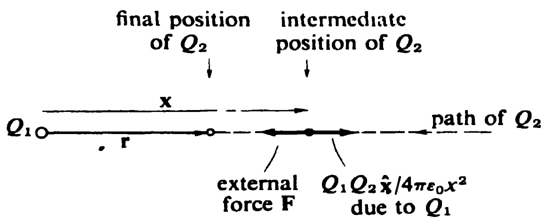


Fig. 2.9. Calculation of the potential energy of two point charges.

chosen to be always along the same direction as r and the external force needed to bring in Q_2 is always equal and opposite to the force repelling it from Q_1 : Q_2 then moves with a constant velocity and no kinetic energy is created. At some intermediate stage when Q_2 is at a distance x from Q_1, the work done by the external force F in a small displacement dx is

$$\mathrm{d}W = F\,\mathrm{d}x = -Q_1Q_2\,\mathrm{d}x/4\pi\varepsilon_0x^2 \qquad (2.8)$$

Hence, the total work done over the whole path, W, or the potential energy, U, is given by

$$W = U = \int_{\infty}^{r} -Q_1Q_2\,\mathrm{d}x/4\pi\varepsilon_0x^2 \qquad (2.9)$$

or

$$U = Q_1Q_2/4\pi\varepsilon_0 r \qquad (2.10)$$

If either charge is not positive, this expression for U will still apply if the appropriate signs for the Q's are inserted when substituting numerical values.

An alternative definition of potential energy does not introduce an external force but uses only the internal ones. The definition would then read 'the work done by the internal forces in taking a system from the given configuration to the standard one'. We should then allow the repulsion alone to take Q_2 from r to infinity. This would interchange the limits of integration in equation (2.9) but would also change the sign in equation (2.8) so that equation (2.10) would be obtained as before. The two definitions of potential energy are equivalent for all systems by a similar argument.

The expression (2.10) contains reference only to the initial configuration which determines the zero and the final one (r), and none to the path taken. We chose a path which made the calculation very simple, but what happens if we choose a different one as in Fig. 2.10?

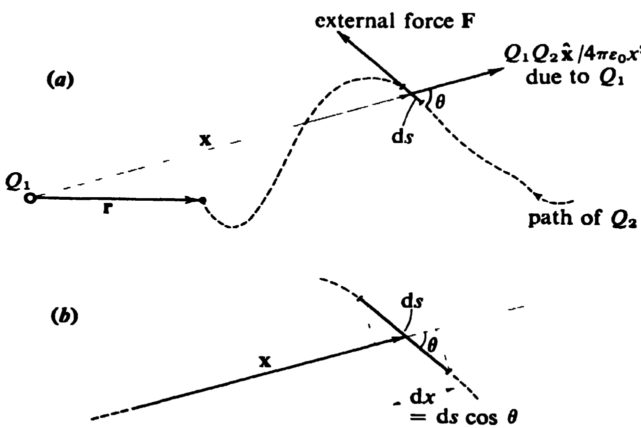


Fig. 2.10. (a) Calculation of potential energy with a general path; (b) detail of an element of the path.

Suppose at some intermediate stage the shortest distance of Q_1 from Q_2 is x so that the internal force is $Q_1Q_2/4\pi\varepsilon_0 x^2$ along **x**. The external force **F** needed is now only $-(Q_1Q_2 \cos\theta)/4\pi\varepsilon_0 x^2$ along ds. Hence the work done by **F** in the increment of path is F ds or $-(Q_1Q_2\,\mathrm{d}s \cos\theta)/4\pi\varepsilon_0 x^2$. Out of this expression we pick ds cos θ and see that it equals dx, the increment in the radial distance from

Q_1. The increment of work, $\mathrm{d}W$, is thus given as before by equation (2.8), and (2.10) follows: U is independent of the path.

Physically, this arises because the force is a central one and work is done only in the radial component of a displacement: any tangential displacement (movement at a constant distance from Q_1) involves no work at all. Only the distance from Q_2 to Q_1 is therefore important.

If Q_2 moves from a distance r' to a distance r from Q_1, the difference in potential energy is

$$U = \frac{Q_1 Q_2}{4\pi\varepsilon_0}\left(\frac{1}{r}-\frac{1}{r'}\right) \tag{2.11}$$

irrespective of the path and of the zero. It should also be clear that, if the path is reversed, the initial and final points are interchanged and the sign of U changes.

The potential energy is *mutual* and does not belong to one charge more than to the other: extension to a collection of charges gives correspondingly the potential energy of the system as a whole.

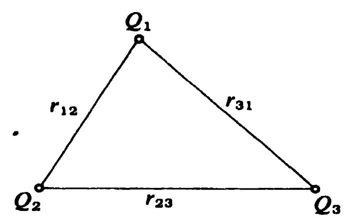


Fig. 2.11. System of three charges.

Thus, because the forces superpose, the charges in Fig. 2.11 have potential energy

$$U = Q_1Q_2/4\pi\varepsilon_0 r_{12}+Q_2Q_3/4\pi\varepsilon_0 r_{23}+Q_3Q_1/4\pi\varepsilon_0 r_{31} \tag{2.12}$$

with respect to the arrangement with all three at infinite distances from each other.

It is most important to understand that the *path-independence* of U in all these cases is a consequence solely of the fact that the forces are central ones. The path-independence has itself the physical consequence that we cannot extract energy from a system of charges by moving one of them in a closed path: if we could, we should have an inexhaustible supply of energy.

2.8 Summary of Chapter 2

We have established that the law of electric force between two charged bodies whose dimensions are negligible compared with their distance apart is

$$\mathbf{F} = Q_1 Q_2 \hat{\mathbf{r}}/4\pi\varepsilon_0 r^2 \tag{2.3}$$

where ε_0 is a constant determined by the units such that

$$\left.\begin{aligned} 4\pi\varepsilon_0 &= 10^{-9}/9 \quad \text{approx. for MKSA units} \\ 4\pi\varepsilon_0 &= 1 \quad\quad\quad\; \text{exactly for CGS e.s.u.} \end{aligned}\right\} \tag{2.6}$$

Verification of the law for all the conditions under which we may wish to use it has not been considered, but we shall assume that it is general, bearing in mind that extrapolations may be made which should be checked by experiment. In particular, its validity for atomic distances is discussed in section 4.3 and for slowly moving charges in section 11.1.

We have seen how to apply the law to collections of charge and to continuous distributions by using the principle of superposition. We have also deduced that the mutual potential energy of two charges Q_1 and Q_2 separated by a distance r is

$$U = Q_1 Q_2/4\pi\varepsilon_0 r \tag{2.10}$$

relative to that of the same charges separated by an infinite distance, and that U is independent of the path by which the charges reach their final positions.

Appendix 2.1 Vector Notation and Addition

Physical quantities specified in terms of a unit by one number only are known as *scalars*: examples are mass and electric charge. Many familiar quantities, however, need three numbers to specify them: examples are velocity, acceleration and forces of many kinds. These are *vectors*, usually defined as quantities with magnitude and direction and represented diagrammatically by a line whose length is proportional to its magnitude and whose direction and sense (indicated by an arrowhead) represent its direction. Figure 2.12 shows that the direction can be unambiguously given with reference to a co-ordinate system by two angles θ and ϕ. The three numbers needed to specify the vector may thus be its magnitude A together with θ and ϕ. Alternatively, the projections of A on to the x-, y- and z-axes, known as the *cartesian components* A_x, A_y and A_z, will also specify the vector.

We shall distinguish vectors by using heavy type thus: **A**. In manuscript it is usual to indicate them by some form of underlining. The magnitude of **A** will be denoted by $|A|$ or, if no ambiguity results, simply by A. Equations involving vectors are often more concisely written using this notation (see equations (2.2)–(2.4)), and

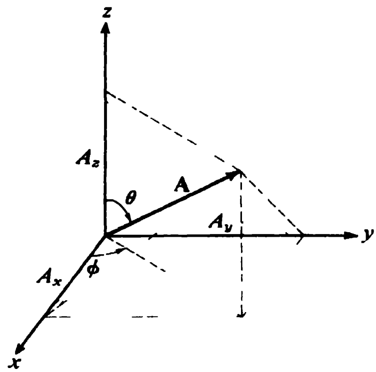


Fig. 2.12. Specification of a vector.

the habit of indicating vectors consistently in equations is in any case advisable so that the directions of quantities are not forgotten.

Vector Addition. The sum of two vectors, $\mathbf{A}+\mathbf{B}$, is found by the parallelogram rule familiar in the case of velocities and forces, and is also given by the triangle as in Fig. 2.13. The sum $\mathbf{A}+\mathbf{A}$ is thus merely a vector in the same direction as **A** but of twice the magnitude

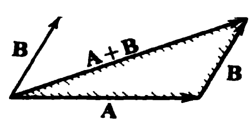


Fig. 2.13. Vector addition by parallelogram or triangle.

and is written 2**A**. Similarly $k\mathbf{A}$ is a vector in the same direction as **A** but k times the magnitude.

This is sometimes used to denote the vector **A** itself as A times a vector of unit magnitude in the direction of **A**. Such a unit vector we shall denote by $\hat{\mathbf{A}}$, so that $\mathbf{A}=A\hat{\mathbf{A}}$ (Fig. 2.14). This has been used in equation (2.3) where **F** is written as $Q_1Q_2/4\pi\varepsilon_0r^2$ times a unit vector $\hat{\mathbf{r}}$.

By a vector $-\mathbf{B}$ we mean one whose sense is opposite to that of $\mathbf{B}$ but is otherwise equal in all respects to $\mathbf{B}$. The subtraction of

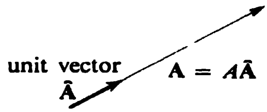


Fig. 2.14. Expression of a vector in terms of a unit vector.

ectors, $\mathbf{A}-\mathbf{B}$, is defined as the addition of $\mathbf{A}$ to $-\mathbf{B}$ as in Fig. 2.15. The associative rule for addition, $\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$, is easily proved.

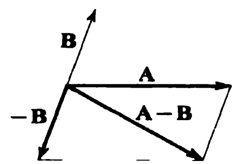


Fig. 2.15. Subtraction of vectors.

Appendix 2.2 Line Integrals

The integral in equation (2.9) is, like all others, essentially a sum, but it is a sum evaluated along a certain line and is known as a line integral. Quantities are often encountered which have a value at every point in a region of space and therefore at every point along a line C in that region between two points A and B (Fig. 2.16). For a small element of the line, form the product of the quantity, say

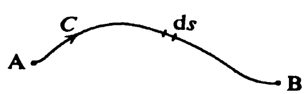


Fig. 2.16. An open path along which the line integral of any function F of position can be evaluated.

$F(s)$, with $\mathrm{d}s$, the length of the element. The line integral of F between A and B along the path C is the sum of all the $F\,\mathrm{d}s$'s from A to B and is written $\underset{C}{\int_{A}^{B}} F\,\mathrm{d}s$. If the quantity is a vector, $\mathbf{F}$, the resolved part of $\mathbf{F}$ along the element, $F\cos\theta$, is multiplied by $\mathrm{d}s$ and

the integral formed. Thus the work done by a force **F** along a path C from A to B is $\int_{\substack{A \\ C}}^{B} F \cos\theta \, ds$.

The evaluation of line integrals is a matter best left to mathematical texts, but it is evident that the result will in general depend both on the positions of A and B and on the path C. In certain cases involving vector quantities, however, the line integral is independent of the path. For instance, we saw in section 2.7 that, if **F** is a central force, then

$$\int_{A}^{B} F \cos\theta \, ds = \int_{r_A}^{r_B} F(r)\, dr = f(r_B) - f(r_A) \qquad (2.13)$$

where r is the radial distance from the centre of force and $f(r)$ is the indefinite integral of $F(r)$, and this depends only on A and B. The physical reason for the path-independence has been discussed in section 2.7. Where path-independence occurs, we choose a path for integration which makes the integration as simple as possible.

Sometimes a line integral round a closed path (as in Fig. 2.17) is

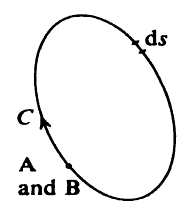


Fig. 2.17. A closed path. A *and* B *in Fig. 2.16 now coincide.*

required: it is conventional to indicate this by the symbol $\oint$. For a function of position F which has a path-independent line integral, $\oint F \, ds$ is zero since A and B in equation (2.13) coincide.

Appendix 2.3 Solid Angles

The solid angle is an extension to three dimensions of a concept familiar enough in two. A *plane* angle in radians (Fig. 2.18) is given by the ratio of s, the arc of the circle centred at O cut off by the arms of the angle, to r, the radius of this circle. For a small angle $d\theta$, the chord and arc differ in length by a second order of smallness and $d\theta$ is given equally well by ds/r and by dl/r.

To define a *solid* angle (Fig. 2.19), a sphere of radius r and centre

at the apex of the angle is constructed and the area cut off on the surface of this sphere by the generators of the angle is S, say The ratio S/r^2 is the solid angle Ω in steradians, the complete solid

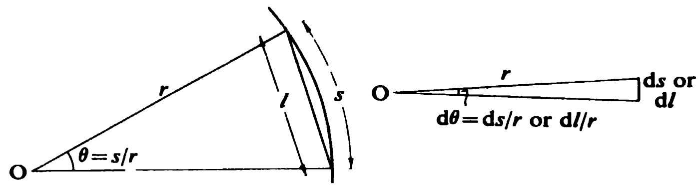


Fig. 2.18. Plane angles.

angle about a point being 4π. A useful result, easily proved, is that the solid angle at the apex of a cone of semi-vertical angle θ is $2\pi(1-\cos\theta)$.

For a small solid angle $d\Omega$, the plane area dA differs from dS by a second order of smallness so that $d\Omega = dS/r^2$ or dA/r^2. If an area dS is not normal to the radius r, then it must be projected into a

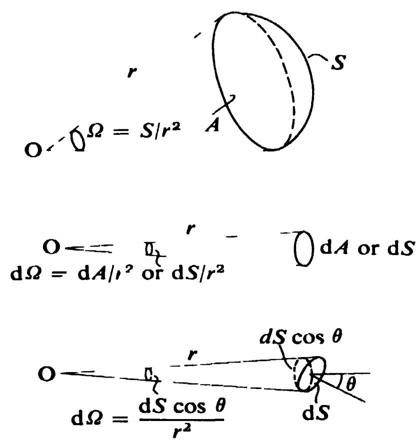


Fig. 2.19. Solid angles.

plane which is normal to r as $dS\cos\theta$: the solid angle subtended by it is then given by

$$d\Omega = (dS\cos\theta)/r^2 \tag{2.14}$$

This is the expression used in section 2.3.

References

Further details of experimental verification of Coulomb's law can be found in Magie (1964) for Coulomb's own work, Maxwell (1904) for Cavendish's and Maxwell's and in Plimpton and Lawton (1936).

Mathematical topics dealt with in the appendices to chapters are treated in more detail and with greater rigour in texts written expressly for science students, such as Stephenson (1961) or Massey and Kestelman (1964). Davis (1961) is particularly recommended for vector methods.

PROBLEMS

SECTION 2.1

2.1 *(*a*) Examine the consequences of mutual forces between point charges not being directed along the joining line.

(*b*) What is wrong with Fig. 2.2?

SECTION 2.3

2.2 Derive the theory behind Coulomb's experimental verification of the inverse square law of *attraction*. The experiment showed that the period of small oscillations of a small charged body at the end of a light horizontal insulating rod pivoted about a vertical axis was proportional to the distance between the body and a fixed charge of opposite sign in the same horizontal plane. The length of the rod was small compared with the distance between charges.

SECTION 2.4

2.3 Estimate the force in tons wt. which would be exerted between two point charges each of 1 C, if they could be placed 1 m apart (1 ton = 1,016 kg).

2.4 Estimate the charge occurring on a rubbed body capable of picking up small pieces of paper.

2.5 Conversion factors between various electric units are easily worked out with a knowledge only of table 1.2 and of the values of $4\pi\epsilon_0$. Find the number, n, of e.s.u. in 1 C by calculating the force between a charge of 1 C and another 1 m from the first in (*a*) MKSA units, (*b*) CGS e.s.u., and equating the results.

SECTION 2.5

2.6 Compare the electric and gravitational forces between two electrons or between two protons a certain distance apart, given that $G = 6{\cdot}67 \times 10^{-11}$ m/kg-s^2. Why do gravitational forces predominate at astronomical distances and electric forces at atomic distances?

SECTION 2.6

2.7 Point charges $+4Q$ and $-Q$ are separated by a distance a. Show that the only positions where a third charge $+Q$ could be in equilibrium are along the line joining the first two, and find any such positions. Is the equilibrium stable or unstable?

2.8 Point charges $+Q$, $-Q$, $+Q$, $-Q$ are situated in order at the corners of a square. Show that any charge placed on a line from the centre of the square perpendicular to its plane is in equilibrium, but is unstable for displacements out of the line.

2.9 A thin rod of length $2l$ possesses a charge λ per unit length. Find the force on a charge Q situated a distance r from the rod along its perpendicular bisector. Check the solution by letting $l \to 0$ and seeing whether the rod behaves as a point charge. What does the force become as $l \to \infty$?

SECTION 2.7

2.10 Show that the solution of problem 2.7 can also be obtained by finding the potential energy U of the system for a general position of $+Q$ along the joining line and determining the positions for which U is a minimum.

2.11 A particle whose charge is $-e$ and whose mass is m is free to move about a fixed point charge $+Ze$. Show that, if the only force acting is the electric one, the particle must move in a plane determined by its initial velocity and position. Show also that, if the particle describes a circle about Ze of radius r and velocity v, the total energy of the system is $-\frac{1}{2}mv^2$.

2.12 If the particle in problem 2.11 can only move in circles for which the angular momentum ($mr^2\omega$ or mvr) is an integral multiple of $h/2\pi$, $nh/2\pi$, show that its total energy can be expressed as $-Z^2e^4m/8\varepsilon_0{}^2h^2n^2$. (Bohr hydrogen-like atom: h is Planck's constant.)

CHAPTER 3

ELECTRIC FIELD STRENGTH AND POTENTIAL DIFFERENCE

Although this chapter and the next two are largely concerned with consequences of Coulomb's law, new concepts make many problems simpler and more vivid. Two of these concepts, electric field strength and potential difference, are introduced in this chapter and used to examine generally the motion of charges in free space, in conductors and in insulators due to electric fields.

3.1 Electric Field Strength

If a stationary charge Q at a point in space experiences a force $\mathbf{F}$ then, provided we ensure that Q in no way affects any charges already present, there is said to be an *electric field* at the point. The strength of this field, denoted by $\mathbf{E}$, is defined as the force per unit positive charge so that

$$\mathbf{E} = \mathbf{F}/Q \qquad \text{(Definition of } \mathbf{E}) \quad (3.1)$$

Two important conditions are included in this definition, the first that Q must not affect charges already present and the second that Q shall be stationary. We discuss these in turn.

In the first place, if Q alters the distribution of charge in any way then the field given by (3.1) will not be the pre-existing one but one depending on Q. As an extreme instance, an uncharged conductor produces no field around it, but the introduction of Q nearby will cause electrostatic induction and a force on Q which, by (3.1), would mean that there *is* a field. This, however, is only going to cause trouble if we use (3.1) to *measure* $\mathbf{E}$: we must then make Q so small that its effect is negligible and (3.1) will become

$$\mathbf{E} = \lim_{Q \to 0} \mathbf{F}/Q \qquad (3.2)$$

The difficulty about this could be that Q cannot be smaller than the electronic charge, but in practice we use neither (3.1) nor (3.2)

directly in the measurement of **E**. We use (3.1) for defining and calculating **E** and in doing so we agree that all charges producing it shall be fixed in position while the force on Q is calculated. This was our procedure in section 2.3: the charges on the sphere were assumed to be fixed and this gave the electric field at the point occupied by Q. *Measurement* of **E** is carried out by finding the potential gradient (section 3.6).

Some authors write (3.2) in the form

$$\mathbf{E} = \lim_{\delta Q \to 0} \delta\mathbf{F}/\delta Q \tag{3.3}$$

and then say that this is just $d\mathbf{F}/dQ$. The $\delta\mathbf{F}$ in (3.3), however, is *not* $\mathbf{F}(Q+\delta Q)-\mathbf{F}(Q)$, but is $\mathbf{F}(\delta Q)$—a different quantity altogether—and $d\mathbf{F}/dQ$ is therefore not a correct expression for **E**. An example in section 4.8 shows this clearly.

We turn to the second proviso in the definition of **E**, that Q shall be stationary. This is necessary to distinguish electric from magnetic fields, for we have seen that currents (which are moving charges) exert magnetic forces on each other, so that if we allowed the definition (3.1) to apply with a moving Q, **E** would include the magnetic force, if any.

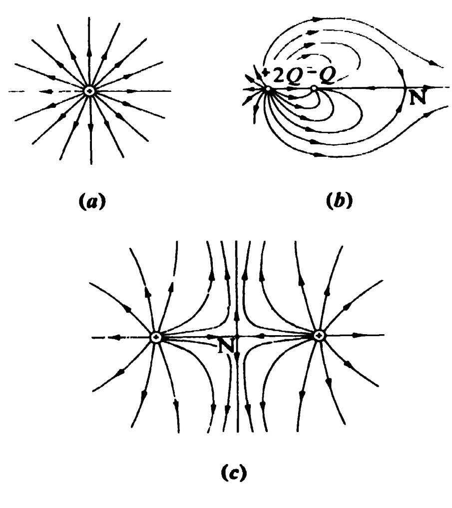


Fig. 3.1. Lines of force due to simple systems of point charges. NN *are neutral points.*

Sources of **E**. While any system producing a force on a stationary charge at a point is a source of electric field, we shall be concerned for the time being only with charges themselves as sources. The electric field produced by stationary charges, known as an *electrostatic field*, is a central force field (from section 2.1) and in this respect we shall see that it is unique among electric fields. For further discussion of this point see section 6.1.

Lines of Force. It is often helpful to visualize electric fields in terms of lines of force—lines in space such that the tangent to them at any point gives the direction of **E** at the point. Figure 3.1 gives some examples which can be calculated from the inverse square law. It is also possible to use the density of lines (the number crossing unit area perpendicular to the field) as an indication of the magnitude of **E**. It is clear from Fig. 3.1 that the stronger fields near the charges are accompanied by a greater density of lines and this idea can be developed more exactly (section 4.1).

Unit of **E**. While the N/C (newton/coulomb) is a correct unit for **E**, it is not the one conventionally used: we return to this in section 3.6.

3.2 Electric Fields due to Charges

We shall consider briefly how to calculate fields produced by given charge distributions using the methods of section 2.6.

For a single point charge Q, the electric field strength at a point a distance $\mathbf{r}$ from it is (Fig. 3.2)

$$\mathbf{E} = \frac{Q}{4\pi\varepsilon_0 r^2}\,\hat{\mathbf{r}} \tag{3.4}$$

giving lines of force as in Fig. 3.1a: equation (3.4) follows from Coulomb's law.

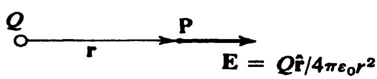


Fig. 3.2. Electric field strength due to a point charge.

For the electric field due to a collection of point charges the principle of superposition yields

$$\mathbf{E} = \frac{Q_1}{4\pi\varepsilon_0 r_1^{\,2}}\,\hat{\mathbf{r}}_1 + \frac{Q_2}{4\pi\varepsilon_0 r_2^{\,2}}\,\hat{\mathbf{r}}_2 + \cdots \tag{3.5}$$

so that the resultant may be found by the addition of vectors, as with forces in section 2 4 (equation (2 5) and Fig 3 3)

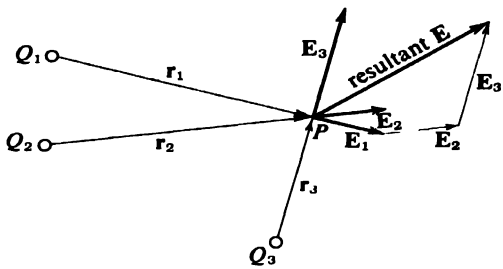


Fig 3 3 Superposition of electric fields.

To calculate the electric field due to a continuous distribution of charge, the method illustrated in section 2 6 for a plane sheet can be used The charge is divided into elements, each exerting a force on the test charge Q calculated from Coulomb's law, these elementary forces are resolved into components, summed by integration and the result divided by Q For instance, the electric field due to a very large plane sheet of charge with uniform surface density σ would be $\sigma/2\varepsilon_0$ by equation (2 7), the lines of force being as in Fig 3 4

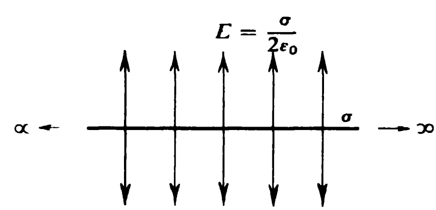


Fig. 3.4 Electric field due to an isolated sheet of charge of surface density σ.

This particular example illustrates a *uniform* electric field, which is one with the same magnitude and direction at all points The charge distribution giving rise to it is, however, hardly a practical one because it consists of a sheet of charge isolated in space In practice it is possible to produce an electric field of any desired uniformity by connecting a battery of cells across two thin plane parallel metallic plates (Fig 3 5) having linear dimensions large compared with their distance apart The battery transfers charge

as usual so that one plate is charged positively and the other negatively, the surface densities being approximately uniform (only for infinite plates could we say that σ was uniform by symmetry). The fields produced by the sheets are each as in Fig. 3.4 with the direction of the lines reversed for the negative one. Figure 3.5

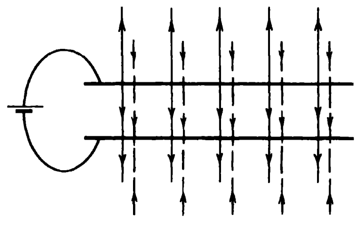


Fig. 3.5. Electric fields due to two parallel charged sheets of equal and opposite charge densities.

shows that the fields cancel in regions outside the plates but reinforce between them, giving an approximately uniform electric field **E** given by

$$E = \sigma/\varepsilon_0 \tag{3.6}$$

in magnitude (Fig. 3.6). This important result will be obtained by less roundabout methods in chapter 4 when we deal with other methods for calculating fields.

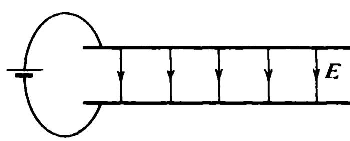


Fig. 3.6. Resultant electric field of Fig. 3.5.

3.3 Charges in Electric Fields—Preliminary

So far we have calculated **E** from known distributions of charge: the other side of the problem is to see what happens when a charge is placed in a pre-existing electric field. It might seem obvious that it would follow from equation (3.1) that a charge Q placed in an electric field of strength **E** would experience a force **F** given by

$$\mathbf{F} = Q\mathbf{E} \tag{3.7}$$

Here, however, we must be very careful indeed that the charge Q

does not alter the pre-existing field for, while (3.1) is only used to calculate **E** theoretically, (3.7) is used to predict what happens in experiments when Q may not be at all small. We shall therefore watch this point carefully: in section 2.3, where we have already effectively used (3.7), the conductors inside the charged spheres were always uncharged and **E** was thus not affected by their presence.

We should also ask whether equation (3.7) will apply if Q is moving. We are already assuming that Coulomb's law still applies when charges are moving and it follows that if **E** is an electrostatic field we are not introducing an extra assumption by retaining (3.7) in those instances.

We have now a means of calculating the force on a charge if we know the electric field in which it is placed but, before considering what will happen to the charge as a result of this force, we shall find it an advantage to introduce the concept of potential difference.

3.4 Electric Potential Difference

In section 2.7 we saw that when one charge was moved in the electric field of another by an external force the work done by this force, or the increase in potential energy, had two important properties: it was independent of the path taken and reversed its sign when the direction of the path was reversed. These properties will hold in the electric field of any system of charges because of superposition.

If we now form the work done *per unit positive charge*, we shall have a quantity which can only depend on the initial and final positions and on the source of the field. This new quantity, unlike **E**, is a scalar and is known as the *potential difference* or *p.d.*

A formal definition of potential difference is required on the lines discussed. We start with a region in which there is an electrostatic field and consider two points A and B joined by any path (Fig. 3.7). A positive charge Q, understood not to affect the pre-existing field, is moved *from A* to B along the path by an external force which has to do work W_{AB} and this increases the potential energy of the system by U_{AB}. The potential difference between B and A is given by

$$V_{AB} = W_{AB}/Q = U_{AB}/Q \quad \text{(Definition of } V_{AB}\text{)} \quad (3.8)$$

or, in words, the work done per unit charge by an external force in taking positive charge from A to B: the point B is at the higher potential when positive work is done.

Strictly, only differences of potential can be defined, as we should expect since they are potential energies per unit charge. We can, however, always choose a zero at some point A, and the potential

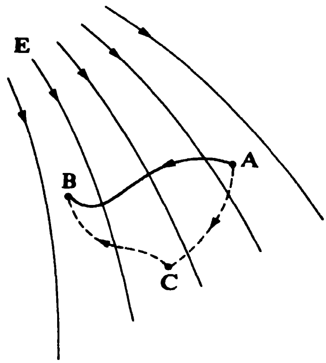


Fig. 3.7. Path AB *for the definition of potential difference. Because the work done is independent of the path,* $W_{AB} = W_{AC} + W_{CB}$.

difference between B and A becomes the *potential* at B with respect to a zero at A:

$$V_B = V_{AB} = W_{AB}/Q = U_B/Q \quad \text{(zero at A)} \quad (3.9)$$

The zero should always be specified in this way when using the term *potential.*

The potentials of two points B and C relative to a zero at A are

$$V_B = W_{AB}/Q; \qquad V_C = W_{AC}/Q = -W_{CA}/Q$$

and hence

$$V_B - V_C = (W_{CA} + W_{AB})/Q = W_{CB}/Q = V_{CB} \qquad (3.10)$$

so that the potential difference between B and C, V_{CB}, can be expressed as the difference of the potentials V_B and V_C measured with respect to any zero.

To summarize, for a *potential* at any point in an electrostatic field, we must choose a zero and evaluate by using (3.9), but if we only require a potential *difference* equation (3.8) can be used independently of a zero (Fig. 3.7).

'*Absolute*' *Potential.* Sometimes the term *absolute potential* is used to mean a potential evaluated with a zero at a point remote

from all other charges (at infinity). We shall not use the term but shall merely specify the zero in the usual way.

Unit of Potential Difference. In the MKSA system this will be the J/C (joule/coulomb) known as the *volt*, symbol V. Thus the potential difference between two points is 1 V when 1 J of work is required to take 1 C of charge from one point to the other.

3.5 Evaluation of Potential Difference

To evaluate a potential difference ($V_B - V_A$) we must find the work done per unit charge by an external force in moving a charge from A to B. In Fig. 3.8, the electric field at some element of the path, ds, is **E**, so that if the charge is to move with constant velocity an external force $\mathbf{F} = -Q\mathbf{E}$ is required. This is the force which does

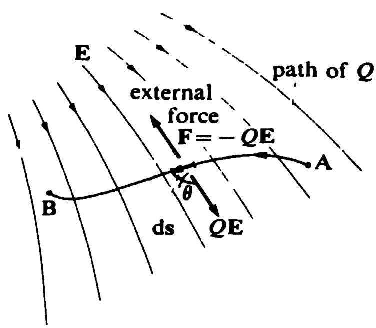


Fig. 3.8. Calculation of potential difference in terms of electric field strength.

the work and in moving Q over the element it will do work $F\,\mathrm{d}s\cos\theta$ or $\mathbf{F}\cdot\mathrm{d}\mathbf{s}$ (see appendix 3.1). Over the whole path the work done is therefore $\int_A^B \mathbf{F}\cdot\mathrm{d}\mathbf{s}$ or $\int_A^B -Q\mathbf{E}\cdot\mathrm{d}\mathbf{s}$. Thus

$$V_B - V_A = \int_A^B -E\cos\theta\,\mathrm{d}s = \int_A^B -\mathbf{E}\cdot\mathrm{d}\mathbf{s} \qquad (3.11)$$

where θ is the angle between **E** and **ds**. Moreover, if A is the zero of potential

$$V_B = \int_A^B -\mathbf{E}\cdot\mathrm{d}\mathbf{s} \qquad \text{(zero at A)} \quad (3.12)$$

In deriving these expressions we have imagined a charge moved

under the action of equal and opposite forces. We must of course make $\mathbf{F}$ infinitesimally larger than $-Q\mathbf{E}$ if Q is to be directed along a chosen path and the extra force $\mathrm{d}\mathbf{F}$ for this purpose will create kinetic energy. The work done by $\mathrm{d}\mathbf{F}$ is small compared with that done by $\mathbf{F}$ and in the limit is negligible: we must agree that the definition implies all this.

Equations (3.11) and (3.12) enable potentials to be determined if the electric fields are known, but they should not be applied without thought—the path may cover several regions in which $\mathbf{E}$ is given by different expressions (see section 4.2). For electrostatic fields the potential difference does not depend on the path, which is therefore usually chosen so as to make the integration simple.

Potentials due to Collections of Charge. The calculation in section 2.7 shows that, if we bring a unit positive charge from infinity to a point P a distance r from a point charge Q,

$$V_{\mathrm{P}} = \int_{\infty}^{r} -Q\,\mathrm{d}x/4\pi\varepsilon_0 x^2 = Q/4\pi\varepsilon_0 r \quad \text{(zero at } \infty) \quad (3.13)$$

using a path always along a line of force for which the angle between $\mathbf{E}$ and $\mathrm{d}\mathbf{x}$ is zero. The potential difference between two points P_1 and P_2 distances r_1 and r_2 from Q will be

$$V_1 - V_2 = \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad (3.14)$$

For a collection of discrete charges the potentials at a point due to the individual charges may be added algebraically since they are scalar quantities and the electric fields superpose. Thus, at a point a distance r_1 from Q_1, r_2 from Q_2, etc.,

$$V = Q_1/4\pi\varepsilon_0 r_1 + Q_2/4\pi\varepsilon_0 r_2 + \cdots \quad \text{(zero at } \infty) \quad (3.15)$$

$$= \frac{1}{4\pi\varepsilon_0}\sum \frac{Q}{r} \quad \text{(zero at } \infty)$$

For continuous distributions of charge, equation (3.13) can be used to find V due to a typical element $\mathrm{d}Q$ and integrating over the whole region occupied by charge. Thus the potentials due to volume, surface and line charges may be written

$$V_\rho = \iiint \frac{\rho\,\mathrm{d}\tau}{4\pi\varepsilon_0 r}; \qquad V_\sigma = \iint \frac{\sigma\,\mathrm{d}S}{4\pi\varepsilon_0 r}; \qquad V_\lambda = \int \frac{\lambda\,\mathrm{d}s}{4\pi\varepsilon_0 r} \quad (3.16)$$

all for a zero at infinity. Alternatively, it may be easier to use a

known expression for $\mathbf{E}$ and to calculate V from (3.12). Care must then be taken in using infinity as a zero when the distribution is not finite (sections 4.4 and 4.5).

Closed Paths. The work done by an external force in taking unit charge round a closed path in an electrostatic field is $\oint -\mathbf{E}\cdot d\mathbf{s}$. If $\mathbf{E}$ is due to a single point charge, this line integral is zero (appendix 2.2 and equation (3.14)). Because the principle of superposition applies, it follows that

$$\oint \mathbf{E}\cdot d\mathbf{s} = 0 \tag{3.17}$$

for *any* electrostatic field $\mathbf{E}$. This condition results only from the fact that the force between charges is a central one and means that *lines of force due to stationary charges cannot close on themselves*: a little thought will show that a closed line of force must violate (3.17).

3.6 Potential Gradient and Field Strength

For a small element of path $d\mathbf{s}$, as in Fig. 3.8, we found that the work done per unit positive charge by the external force was $-\mathbf{E}\cdot d\mathbf{s}$. This is an increment of the potential difference dV, so that

$$dV = -\mathbf{E}\cdot d\mathbf{s} \tag{3.18}$$

But the scalar product is equal to $E_s\, ds$ (appendix 3.1) where E_s is the component of $\mathbf{E}$ in the direction $\mathbf{s}$, and hence

$$E_s = -\partial V/\partial s \tag{3.19}$$

where partial derivatives occur because V in general depends on three co-ordinates (appendix 3.2). In words, (3.19) states that the resolved part of $\mathbf{E}$ in any direction is equal to the negative rate of change of V in that direction or to the negative *potential gradient*. In particular, if V is expressed in terms of cartesian co-ordinates x, y, z,

$$E_x = -\frac{\partial V}{\partial x}; \qquad E_y = -\frac{\partial V}{\partial y}; \qquad E_z = -\frac{\partial V}{\partial z} \tag{3.20}$$

If V can be specified in terms of plane polar co-ordinates (r, θ) only, then the radial and tangential components of $\mathbf{E}$ are

$$E_r = -\frac{\partial V}{\partial r}; \qquad E_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta} \tag{3.21}$$

(see appendix 3.3). The relations (3.19)–(3.21) can be used to obtain

E from V when it is easier to calculate the latter independently. Moreover, electric fields are measured by finding potential gradients rather than forces on charges, so the unit of **E** in the MKSA system is conveniently the V/m: this is the same as the N/C. In practice, fields may be quoted in V/cm, kV/mm, etc., depending on their magnitudes.

As an example, the uniform field between parallel plates 2 cm apart in Fig. 3.9 is 2 V/cm or 200 V/m when the potential difference

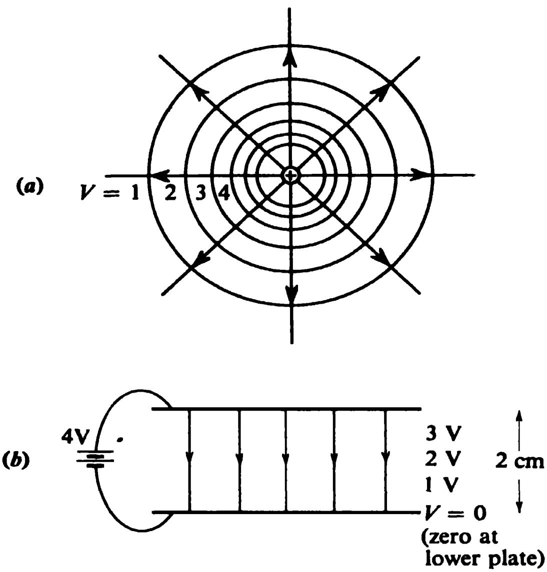


Fig. 3.9. Lines of force and cross-sections through equipotential surfaces for (a) a point charge, (b) a uniform field.

across the plates is 4 V. In general, for a uniform field in the x-direction $dV/dx = -E$ and, since E is constant,

$$V = C - Ex \tag{3.22}$$

where C is a constant depending on the point chosen for the zero of V.

Equipotentials and Lines of Force. Equipotential surfaces are those connecting points at the same potential: in a diagram we can only show two-dimensional cross-sections. Figure 3.9a shows equipotentials of a point charge at equal intervals of potential.

Along any direction tangential to an equipotential there can be no change of V and hence, by (3.19), there can be no component of **E** in this direction. It follows that the lines of force are always normal to the equipotentials (as in Fig. 3.9). Considering various directions between the normal and the tangent to the surfaces as in Fig. 3.10, we see that the rate of change of V is greatest in the

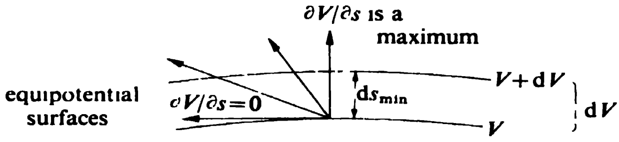


Fig. 3.10. dV is the same for all paths between the surfaces, but ds is a minimum for the direction normal to both and hence $\partial V/\partial s$ is a maximum.

direction of **E**. It is also clear that equipotentials spaced more closely mean a greater maximum rate of change of V and hence a greater **E**.

The equipotentials are analogous to the contours on a map where

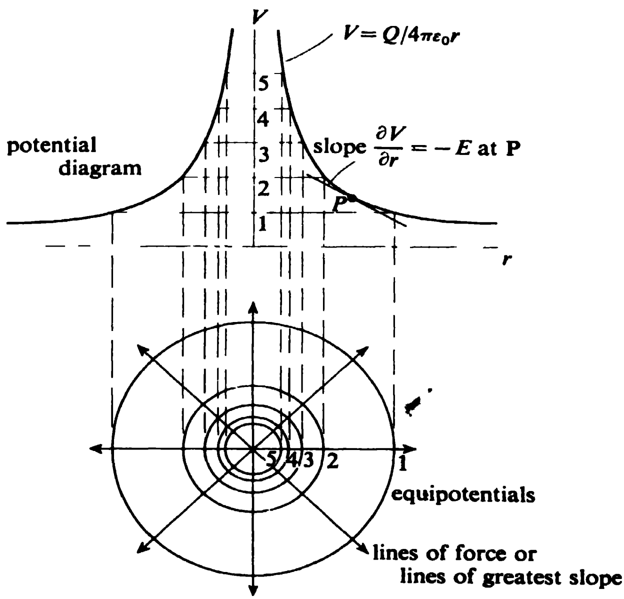


Fig. 3.11. Potential diagram for a positive point charge, showing the analogy between equipotentials and contours.

lines of greatest slope are at right angles to the contours, while the slopes themselves are greatest where the contours are closest. We can carry the analogy further by drawing a cross-section through Fig. 3.9 as if it were a hill and Fig. 3.11 shows the result. In the latter figure the curve of V against r, which appears as the profile of the hill, is known as a *potential diagram*. We shall have occasion later to use such diagrams in which the vertical axis may represent either potential as in Fig. 3.11, or potential energy (see section 4.3 and appendix 4.2).

3.7 The Action of Electric Fields on Charged Particles *in Vacuo*

We now return to the problem we left in section 3.3: what happens to charges in regions where electric fields exist? In this section we deal with charges in free space and in the next with conductors and insulators.

In free space we make the assumption either that the charge is small enough not to affect the field in which it is placed or that the field can be maintained constant by its sources. If no other forces act on a charge Q in an electric field $\mathbf{E}$, its equation of motion is

$$Q\mathbf{E} = m\mathbf{a} \tag{3.23}$$

so that the acceleration $\mathbf{a}$ is $Q\mathbf{E}/m$, where m is the mass of the particle carrying the charge. If $\mathbf{E}$ is known as a function of position this becomes a problem similar to many encountered in particle mechanics: the path can be calculated.

Now consider the motion between two points in terms of potential. Because a positively charged particle accelerates in the direction of $\mathbf{E}$ and because the potential falls along a line of force in the same direction, a positive charge tends to move from regions of higher potential to those of lower and a negative charge will do the opposite. If a charge is moved as in Fig. 3.12, the potential fall from A to B is $V_A - V_B$ and from (3.8) the potential *energy* U_{AB} lost by Q is

$$U_{AB} = Q(V_A - V_B) \tag{3.24}$$

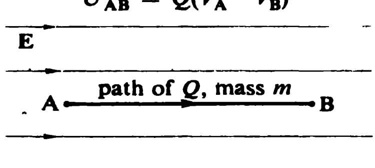


Fig. 3.12. Acceleration of a charged particle in an electric field.

As with a particle falling under gravity this loss in potential energy is compensated by a gain in kinetic energy: for, since $V_A - V_B = \int_B^A -E\,dr = \int_A^B +E\,dr$ for a charge moving along a line of force,

$$U_{AB} = \int_A^B QE\,dr = \int_A^B ma\,dr \quad \text{by (3.23)}$$

$$= \int_A^B mv\,dv \quad \text{writing } a = v\,dv/dr$$

$$= \tfrac{1}{2}mv_B{}^2 - \tfrac{1}{2}mv_A{}^2 \qquad (3.25)^*$$

Although (3.25) was derived for a charge moving along a line of force, the velocity perpendicular to **E** will be unchanged and hence the relation applies to any trajectory.

In atomic and nuclear physics charged particles of high kinetic energy are produced by acceleration in electric fields (section 11.4) and rather than quote the energies of such particles in joules, it is more convenient to use as a unit the kinetic energy gained by an electron in falling *in vacuo* through a potential difference of 1 V, known as the *electron-volt*, symbol eV. For higher energies 1 MeV and 1 GeV (see glossary of symbols) are often used as units although none of these is an MKSA unit. The latter remains the joule, related to the eV by 1 eV $= 1{\cdot}6 \times 10^{-19}$ J. Chapter 11 deals with motion in electric fields in more detail.

3.8 Conductors and Insulators in Electric Fields

Conductors. We know already that a conductor contains charges free to move within it so that, if an electric field is established in conducting material, charges will move as long as the field exists.

If the conductor is isolated, Fig. 3.13 illustrates what will happen when it is placed in a field. As soon as **E** is established, one part of the conductor is at a higher potential than another and any free positive charges will move in the direction of **E** (and any negative charges in the opposite direction). Whatever the sign of the moving charges, the result is the same: charges reach the surface of the conductor and can go no further. They collect and produce a field within the conductor which opposes the applied field. This process continues until, within the material, there is no resultant field. Static conditions will then again prevail.

* The last integration assumes that v is always much smaller than c, the velocity of light, and m is therefore constant (see chapter 11).

We have thus accounted for electrostatic induction mentioned in section 1.1 and in addition have shown that *a conductor carrying only static charge can have no electric field within its material*, and hence that throughout the volume of such a conductor there is no

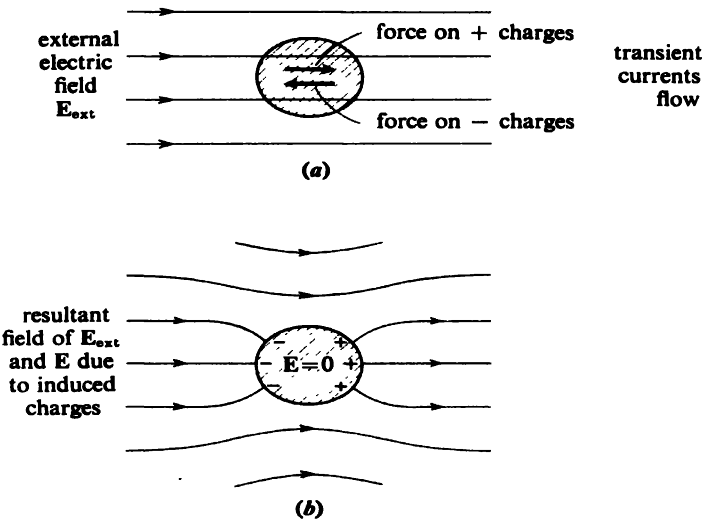


Fig. 3.13. ***An isolated conductor in an electric field. (a) Initial state; (b) final state: electrostatic induction.***

potential difference. *A conductor carrying static charge is an equipotential volume and its surface an equipotential surface.* A similar argument shows that the electric field from the surface is at right angles to it at all points.

If any number of conductors at different potentials are connected by conducting wires, the potential gradients in the wires will cause currents to flow until the potentials are the same: *connecting together charged conductors equalizes their potentials* (Fig. 3.14).

Sources of E.m.f. and of Steady Currents. While the potentials of the two conductors of Fig. 3.14 are equalizing, a transient current flows in the connecting wire as we saw in section 1.3. If a device could be arranged to take the charge from one conductor back to the other as fast as it arrives, the electrostatic field and potential difference between them would be maintained and a steady current set up in the connecting wire (Fig. 3.15). Such a device, known as a

source of electromotive force or e.m.f., we shall examine in more detail in section 6.1, merely noting for the moment that a voltaic cell is one example of such a source, that these sources cause charge to

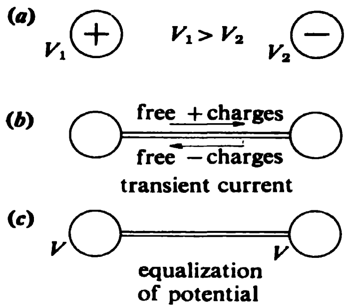


Fig. 3.14. Interconnection of conductors equalizes their potentials.

flow round a complete circuit or closed path, and that, although the potential energy lost by a charge Q in the wire is $Q(V_1 - V_2)$ as before, no kinetic energy is now gained but heat is produced instead.

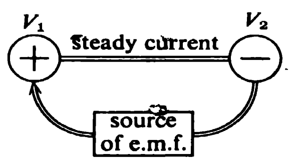


Fig. 3.15. Action of a source of e.m.f. By conveying charges through itself it maintains the potential difference $V_1 - V_2$ and a steady current in the wire. The arrow indicates the direction in which positive charges move.

Insulators. A perfect insulator contains no free charges but we know from section 1.1 that a small movement of + and − charge in opposite directions must occur in an electric field. This displacement of charge, which disappears on removal of the field, is known as *polarization*; further consideration of this is left until chapter 13.

3.9 Determination of the Electronic Charge

We can now look in more detail at the evidence for the existence of a lower limit to the amount of electric charge found in nature (the

electronic charge mentioned in section 1.5) and see how it is measured.

It is now generally recognized that the most accurate determination of e is indirect (e.g. measuring the faraday F and Avogadro's number N_A independently and using $e = F/N_A$), but the oil-drop method whose development was largely due to R. A. Millikan is a convincing direct experiment and in the modification of Hopper and Laby (1941) gave a value with an estimated error of only 0·03 %.

The principle of the method is to observe the velocity of minute droplets of apiezon oil and castor oil falling through air, first in the absence of an electric field and then with a field applied horizontally. The drops were produced in a box vertically above the apparatus by the simple device of pushing the hairs of a wire brush with some oil on their ends, and then releasing them (Fig. 3.16). A copious

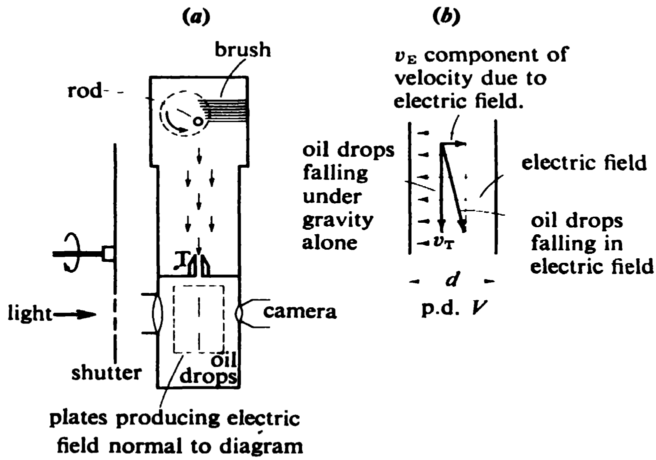


Fig. 3.16. Hopper and Laby's oil-drop experiment for the determination of the electronic charge.

supply of drops fell, and some passed through a tube T into a region where they were illuminated and photographed. A shutter consisting of a rotating disc with apertures allowed the light to pass 25 times per second for 1/1,500 second duration, thus producing on the photographic plate a series of images of the drop as it fell. When the electric field was applied, the direction of fall changed if the drop was charged from ions which are always naturally present in the atmosphere.

The advantage of applying the field horizontally rather than vertically as in Millikan's original method is that convection currents in the air, changes in the charge on the drop and other disturbing factors are easily detected.

To understand what measurements are needed in this method, it should be realized that when a body falls through a viscous medium, like the oil-drop through air, the viscous drag opposing the motion increases with velocity: for a spherical body of radius a falling with a velocity v through a medium whose coefficient of viscosity is η, the opposing force is given by Stokes's law:

$$\text{Viscous drag} = 6\pi\eta a v$$

As the body falls and accelerates, the drag increases and the acceleration is reduced. The velocity at which the weight and drag are equal (so that the speed is constant) is known as the *terminal velocity*. Hopper and Laby showed that their spherical drops of radii about 5 μ took less than 0·003 s to be within 0·01 % of their terminal velocity, so that all those photographed would be travelling with speeds which were constant to within this limit.

Before the electric field is applied, let a drop of radius a and density ρ_0 be falling vertically with a terminal velocity v_T in air of density ρ_a so that, equating weight and viscous drag,

$$4\pi a^3(\rho_0-\rho_a)g/3 = 6\pi\eta a v_T$$

or

$$a^2 = 9\eta v_T/2(\rho_0-\rho_a)g \tag{3.26}$$

Now suppose that the drop possesses a charge Q and that an electric field E is applied, producing a terminal velocity v_E horizontally. Then

$$QE = 6\pi\eta a v_E \tag{3.27}$$

Since a is known from (3.26), Q can be determined in terms of the field E $(=V/d)$, the viscosity of air η, the densities, the velocities v_T and v_E obtained from the photographs, and g.

The values of Q obtained for 16 different charged drops were all multiples of a common value, the multiples being 233, 276, 150, 82, 57, 105, 40, 66, 110, 54, 141, 35, 55, 69, 133 and 231. The conclusion is that the common value is e, the electronic charge.

The most accurate value for e to date is that quoted by Cohen (1955):

$$e = (1{\cdot}60206 \pm 0{\cdot}00003)\times 10^{-19}\ \text{C}$$

while King (1960) has shown also that the charges on the electron and proton do not differ in magnitude by more than 1 part in 10^{20}.

3.10 Summary of Chapter 3

The introduction of electric field strength $\mathbf{E}$ apparently does nothing more than break Coulomb's law into two parts:

$$\begin{tikzcd} & \mathbf{F} = \dfrac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}\,\hat{\mathbf{r}} \arrow[dl] \arrow[dr] & \\ \mathbf{F} = Q\mathbf{E} & & \mathbf{E}_{\text{pt charge}} = \dfrac{Q}{4\pi\varepsilon_0 r^2}\,\hat{\mathbf{r}} \end{tikzcd} \qquad \text{(3.7) and (3.4)}$$

yet even in this chapter we see how some electrostatic phenomena are simplified by thinking of them in terms of electric fields and, because of the way $\mathbf{E}$ has been defined, we realize that there will be sources of $\mathbf{E}$ other than static charges.

Electric potential difference is slightly different. In defining it as the work done by an external force in taking unit charge from one point to another, we have that

$$V_B - V_A = \int_A^B -\mathbf{E}\cdot d\mathbf{s} \tag{3.11}$$

and

$$E_s = -\partial V/\partial s \tag{3.19}$$

But here, only if $\mathbf{E}$ is an *electrostatic* field can we guarantee that $V_B - V_A$ will be independent of the path between B and A and will have a unique value, summed up by the property that $\oint \mathbf{E}\cdot d\mathbf{s} = 0$. For other sources of $\mathbf{E}$ we may find that the idea of potential difference is not a useful one and that $\oint \mathbf{E}\cdot d\mathbf{s} \neq 0$.

Both $\mathbf{E}$ and V are useful quantities because they are independent of the test charges used to define them but V, being a scalar, is often the easier to use.

The calculation of $\mathbf{E}$ and V due to collections of charge may be carried out using one of two methods:

1. Using equation (3.5) to calculate $\mathbf{E}$ (if necessary by integration), and (3.11) to obtain V from $\mathbf{E}$;
2. using (3.15) or (3.16) to calculate V (zero at infinity), and (3.19) to obtain $\mathbf{E}$ from V.

Clearly, method 1 is better when $\mathbf{E}$ is initially easier to find than V.

In an electric field (*a*) charges *in vacuo* are accelerated, gaining kinetic energy, (*b*) conductors undergo electrostatic induction, but

no field can exist in the material of the conductor when static charges only are carried, (*c*) an insulator polarizes. Finally, we have direct evidence that electric charge exists in discrete amounts as we surmised in chapter 1.

Appendix 3.1 The Scalar Product of Two Vectors

Although we have attached a meaning to the addition and subtraction of vectors and to the multiplication of a vector by a number, there is no obvious meaning to be given to the *product* of two vectors. We find in physical applications, however, that there are two functions of pairs of vectors which are common and useful: one is itself a vector and this we leave till later (appendix 7.2); the other is a scalar known as the *scalar product.*

Two vectors **A** and **B** (Fig. 3.17) with an angle θ between them are said to have a scalar product $\mathbf{A}\cdot\mathbf{B}$ (read as 'A dot B') defined by

$$\mathbf{A}\cdot\mathbf{B} = |A||B| \cos\theta \tag{3.28}$$

so that it is the product of the magnitude of either vector and the projection on it of the other.

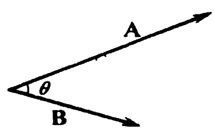


Fig. 3.17. Scalar product $\mathbf{A}\cdot\mathbf{B} = AB\cos\theta$.

Scalar products occur whenever one vector has to be resolved in the direction of another: for example, the work done by a force **F** when its point of application moves a distance **s** is $Fs\cos\theta$ (Fig. 3.18) or $\mathbf{F}\cdot\mathbf{s}$; and if $\hat{\mathbf{s}}$ is a unit vector in a direction **s** then any vector **A** has a component $A_s = \mathbf{A}\cdot\hat{\mathbf{s}}$

The following properties are easily derived from the definition: (*a*) $\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}$, (*b*) $\mathbf{A}\cdot\mathbf{B} = 0$ if **A** is perpendicular to **B**, (*c*) $\mathbf{A}\cdot\mathbf{B} = AB$ if **A** is parallel to **B** (see also problem 3.13), (*d*) $\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C}$, (*e*) $(k\mathbf{A})\cdot\mathbf{B} = k(\mathbf{A}\cdot\mathbf{B})$.

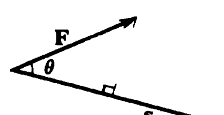


Fig. 3.18. The work done by **F** *displaced through* **s** *is* $Fs\cos\theta$ *or* $\mathbf{F}\cdot\mathbf{s}$.

Appendix 3.2 Partial Differentiation

When a quantity Z is a well-behaved* function of one other quantity x, it can be represented by a curve $Z=f(x)$ as in Fig. 3.19: dZ/dx can be evaluated at any point such as P, and is equal to the slope of the tangent to the curve at that point.

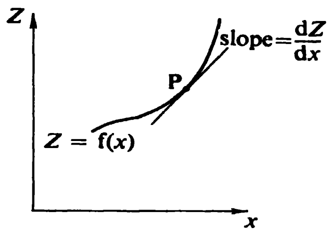


Fig. 3.19. The line $Z=f(x)$.

If, however, Z is a function of two quantities, x and y, then $Z=f(x, y)$ is a *surface* in x, y, Z co-ordinates, part of which is drawn in Fig. 3.20. In this case, dZ/dx can have an infinite number of values at a point P depending on how y is allowed to vary. An example is the rate of change of height of a hill with distance measured north-south: the result depends entirely on how much east-west displacement is allowed.

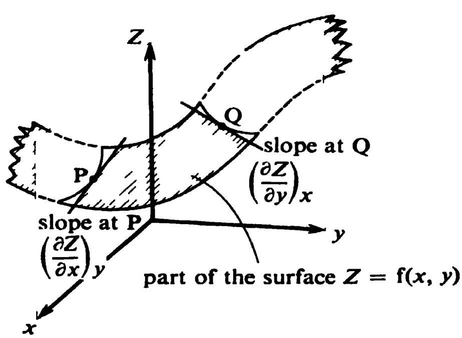


Fig. 3.20. The surface $Z=f(x, y)$.

* We usually assume *at least* that the function and its first derivative are single-valued, continuous functions of x

When we have $Z = f(x, y)$ we commonly use two rates of change—that of Z with x when y is held constant, denoted by $(\partial Z/\partial x)_y$, and of Z with y when x is held constant, $(\partial Z/\partial y)_x$. In Fig. 3.20 these correspond to the slopes at P and Q of sections of the surface cut by xZ- and yZ-planes respectively. They are known as partial differential coefficients or partial derivatives.

We often wish to know the increment in Z when both x and y vary by dx and dy respectively. Since the increment due to dx alone is $(\partial Z/\partial x)_y\, dx$ and that due to dy alone is $(\partial Z/\partial y)_x\, dy$, the total differential dZ is

$$dZ = \left(\frac{\partial Z}{\partial x}\right)_y dx + \left(\frac{\partial Z}{\partial y}\right)_x dy \tag{3.29}$$

For functions of more than two variables, say $Z = f(x, y, z)$, the use of $(\partial Z/\partial x)$ implies that all the other variables are to be held constant. For instance, if $Z = x^2y + 2y^2 + 3z^2$, then $\partial Z/\partial x = 2xy$; similarly, if $Z = xy + yz$ then $\partial Z/\partial x = y$. The function Z cannot now be drawn but the total increment in Z due to increments in x, y and z is, by an extension of (3.29),

$$dZ = \left(\frac{\partial Z}{\partial x}\right) dx + \left(\frac{\partial Z}{\partial y}\right) dy + \left(\frac{\partial Z}{\partial z}\right) dz \tag{3.30}$$

Appendix 3.3 Polar Co-ordinates

Instead of specifying the position of a point in a plane with cartesian co-ordinates x and y, the plane polar co-ordinates (r, θ) may be used (Fig. 3.21). For a point P the relation between the two is that $x = r\cos\theta$, $y = r\sin\theta$; or $r = (x^2 + y^2)^{1/2}$, $\theta = \tan^{-1} y/x$. It is sometimes simpler to express quantities like **E** and V in terms

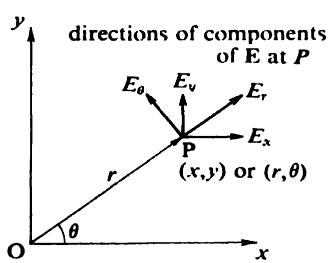


***Fig. 3.21.** Plane polar co-ordinates. The electric field at P has cartesian components E_x and E_y, and polar co-ordinates E_r and E_θ.*

of (r, θ) than of (x, y), just as it is simpler to write the field due to a point charge in terms of r only than in terms of (x, y, z).

The components of a vector in plane polar co-ordinates are the radial and tangential, denoted by E_r and E_θ for **E** as in Fig. 3.21. If a potential V is known in terms of r and θ, application of equation (3.18) gives immediately $E_r = -\partial V/\partial r$ but *not* $E_\theta = -\partial V/\partial\theta$ which is clearly dimensionally incorrect. Figure 3.22 shows that the

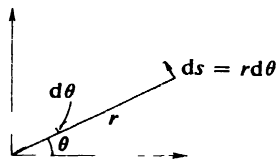


Fig. 3.22. An element of displacement perpendicular to the radial co-ordinate.

element of length ds in a tangential direction is $r\,d\theta$, so that

$E_\theta = -\dfrac{1}{r}\dfrac{\partial V}{\partial\theta}$ as in equations (3.21).

In three dimensions, a z-axis can be added perpendicular to the $r\theta$-plane and a position specified by *cylindrical* polar co-ordinates (r, θ, z): **E** will then have a third component $E_z = -\partial V/\partial z$. Another three-dimensional co-ordinate system is that employing *spherical* polar co-ordinates (r, θ, ϕ) which correspond to the (A, θ, ϕ) of Fig. 2.12. Note that r in the latter system is the distance from the *origin*, while in the cylindrical system it is the perpendicular distance from the z-axis.

PROBLEMS

SECTION 3.2

3.1 Point charges $+Q$ and $-Q$ are separated by a distance $2l$, and O is the point midway between them. Find the electric field strength at P, a distance r from O, (*a*) when OP is along the line joining the charges and $r > l$, (*b*) when OP is perpendicular to the line joining the charges. Show that when $r \gg l$ both fields are nearly proportional to Ql/r^3. (The charges form an electric *dipole* and P is in the *end-on* and *broadside* positions respectively —see section 4.6.)

3.2 A thin rod of length $2l$ is uniformly charged with λ per unit length. Find the electric field strength at a point on the line of the rod, beyond its ends and a distance a from its midpoint.

3.3 A charge Q is spread uniformly round a thin wire bent into a circle centre O radius a. Show that the electric field strength at a point P on the axis of the circle is $(Q \cos \theta)/4\pi\varepsilon_0 r^2$ where r is the distance of P from any point on the wire and θ is the angle between r and the axis. Hence show that the work done in taking unit charge from O to infinity is $Q/4\pi\varepsilon_0 a$.

SECTIONS 3.4–3.6

3.4 Find the potential (zero at infinity) at the point on the line of the charged rod of problem 3.2.

3.5 According to the calculation in section 2.6, the electric field strength on the axis of a circular sheet of charge is $(1 - a/(a^2 + b^2)^{1/2})\sigma/2\varepsilon_0$. Derive this by evaluating the potential at the point and using (3.19).

3.6 Show that the potential energy of a finite collection of charges can be expressed as $\sum_i \frac{1}{2}Q_i V_i$ where V_i is the potential at the ith charge due to all the other charges.

3.7 Sketch potential diagrams for (*a*) two positive charges at various distances apart, (*b*) a positive and a negative charge. Infer that a positive charge raises the potential everywhere and a negative charge lowers it. Sketch a *potential energy* diagram for a row of positive charges in the field of which an electron moves: this is a simple model of a metallic conductor.

3.8 Show that the equations of lines of force in the xy-plane are given by the solutions of $\mathrm{d}y/\mathrm{d}x = E_y/E_x$; and in the $r\theta$-plane by the solutions of $\mathrm{d}r/\mathrm{d}\theta = rE_r/E_\theta$. Find the equations for a point charge at the origin in both cartesian and polar co-ordinates and check that they give the lines of Fig. 3.1a.

SECTION 3.7

3.9 An electron starts from the cathode of a thermionic valve with negligible velocity. What are its kinetic energy and velocity at the anode if the potential difference between electrodes is 100 V? How many electrons per unit volume at the anode are there if 1 mA is collected by every square cm of electrode? Explain how it is that the current in such an electron beam is constant when v in equation (1.15) is increasing.

3.10 Find the relation between velocity and potential difference using classical mechanics for (*a*) electrons, (*b*) protons accelerated *in vacuo*. For what potential differences do the velocities become 1/10 of the velocity of light? (See chapter 11 for relativistic expressions.)

3.11 A beam of electrons each with velocity v, charge e and mass m is projected into the space between the plates of Fig. 3.9 at right angles to the lines of force. If the length of the plates is l, their distance apart x and their potential difference V, find the angle at which the beam emerges assuming that the field is uniform to the ends of the plates and does not extend beyond them. Show that the linear deflection of the beam at a large distance from the plates is proportional to V. (Electrostatic deflection in a cathode-ray tube.)

3.12 How long does an electron take to travel between two parallel plates separated by 1 cm between which a potential difference of 100 V is maintained? Assume that the initial velocity at one plate is zero. (This is the *transit time* between electrodes.)

APPENDIX 3.1

3.13 If **i**, **j**, **k** are unit vectors along the x-, y-, z-directions respectively, any vector **A** can be written as $\mathbf{i}A_x+\mathbf{j}A_y+\mathbf{k}A_z$. Show that (*a*) $\mathbf{i}\cdot\mathbf{i}=1$, $\mathbf{i}\cdot\mathbf{j}=0$ etc., (*b*) $\mathbf{A}\cdot\mathbf{B}=A_xB_x+A_yB_y+A_zB_z$, (*c*) $\mathrm{d}(\mathbf{A}\cdot\mathbf{B})/\mathrm{d}t=\mathbf{A}\cdot\mathrm{d}\mathbf{B}/\mathrm{d}t+\mathbf{B}\cdot\mathrm{d}\mathbf{A}/\mathrm{d}t$. ($\mathrm{d}\mathbf{A}/\mathrm{d}t$ is defined as $\lim_{\delta t\to 0}[\mathbf{A}(t+\delta t)-\mathbf{A}(t)]/\delta t$.)

APPENDIX 3.2

3.14 If $r^2=x^2+y^2$ and $\theta=\tan^{-1} y/x$, find the useful partial derivatives $\partial r/\partial x$ and $\partial\theta/\partial x$. By writing $x=r\cos\theta$, $y=r\sin\theta$, find also $\partial x/\partial r$ and $\partial x/\partial\theta$.

APPENDIX 3.3

*3.15 Show that in spherical polar co-ordinates the element of displacement normal to the $r\theta$-plane is $r\sin\theta\,\mathrm{d}\phi$ and hence that the components of **E** are $E_r=-\partial V/\partial r$, $E_\theta=-(1/r)\partial V/\partial\theta$, $E_\phi=-(1/r\sin\theta)\,\partial V/\partial\phi$. (See Fig. 2.12 for angles.)

CHAPTER 4

FURTHER ELECTROSTATIC METHODS AND PROBLEMS

Our concern in the first three chapters has been to establish some basic laws and concepts. We are now in a position to use these to solve some electrostatic problems of particular importance both to electromagnetism and to molecular, atomic and nuclear physics. No new laws will be introduced and no new concepts except that of flux: the whole of the chapter is essentially a set of deductions from Coulomb's law.

Gauss's theorem, with which we start, is an alternative form of the inverse square law and should not be confused with a theorem in vector analysis of the same name. We shall use it to deduce some very general results and also to find the fields of some highly symmetrical distributions of charge. Section 4.2, dealing with Poisson's and Laplace's equations, may be omitted at first reading but the section is important for later work: its methods are typical of those used to put laws into differential forms which apply at a point.

The electric dipole occurs frequently in modern physics and warrants the detailed treatment given to it in section 4.6: most of the results are equally applicable to magnetic dipoles and will be needed again in chapter 7. The quadrupole and higher multipoles are not normally introduced at this level but the brief outline of their meaning and properties in section 4.7 are essential to the understanding of some aspects of atomic spectra and nuclear properties.

Finally, since all the problems considered up to section 4.7 are ones in which charge distributions are *given*, it is desirable to look at a simple example of a method which allows distributions to be calculated, the method of images (section 4.8).

4.1 Gauss's Theorem

In section 3.1 it was promised that the idea of using the density of lines of force as a measure of **E** would be developed more precisely.

In elementary electricity this is done by letting the magnitude of **E** at any point equal the number of lines of force per unit area crossing an area perpendicular to **E** at that point. Taking now a point charge Q at the centre of a sphere of radius r, we find the value of E at all points on the surface to be $Q/4\pi\varepsilon_0 r^2$ and, since the area of the surface is $4\pi r^2$, the convention adopted means that Q/ε_0 lines of force leave a charge Q.

We shall put this idea into more formal terms by using, instead of the density of lines, the flux of the vector **E** over an area (explained in appendix 4.1). A point charge Q is enclosed by an imaginary surface S (as in Fig. 4.1), known as a *Gaussian surface*. An element

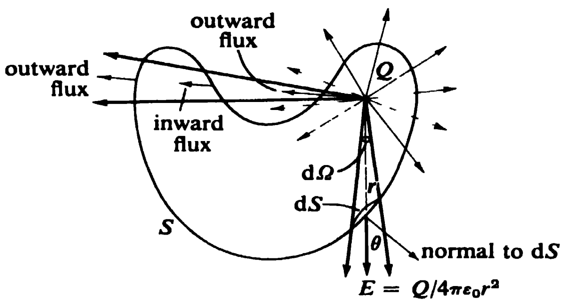


Fig. 4.1. Derivation of Gauss's theorem.

dS of this surface is a distance r from Q and the electric field strength E at dS is thus $Q/4\pi\varepsilon_0 r^2$ in the direction of r. Hence,

$$\text{Flux of } \mathbf{E} \text{ across } dS = \frac{Q\, dS \cos\theta}{4\pi\varepsilon_0 r^2}$$

$$= \frac{Q\, d\Omega}{4\pi\varepsilon_0}$$

where $d\Omega$ is the solid angle subtended by dS at Q. The total flux of **E** over S is the sum of all these elementary fluxes which is simply $Q/4\pi\varepsilon_0$ times the total solid angle at Q, 4π. Hence,

$$\text{total outward flux of } \mathbf{E} \text{ across } S = Q/\varepsilon_0 \tag{4.1}$$

where the sign of the charge is incorporated in Q and inward flux is negative.

Any charges outside S contribute as much inward as outward flux while the shape of S makes no difference to these results as can be

seen from Fig. 4.1. Moreover, because electric fields superpose, a number of charges inside S give their own fluxes each obeying (4.1); these fluxes add, giving a general theorem that *the outward flux of* **E** *over any closed surface is equal to the algebraic sum of the enclosed charges divided by* ε_0, or

$$\oint\!\!\oint \mathbf{E}\cdot d\mathbf{S} = \sum Q/\varepsilon_0 \tag{4.2}$$

which is Gauss's theorem*. Notice that, once **E** is defined, this is a deduction from Coulomb's law and the principle of superposition.

General Deductions from Gauss's Theorem We know already (section 3.8) that in the material of a conductor carrying static charges only, **E** is everywhere zero. Any Gaussian surface drawn entirely within conducting material can therefore have no flux of **E** crossing it and must contain zero charge. A conductor with no interior surface (Fig. 4.2a) therefore *carries all its excess charge on*

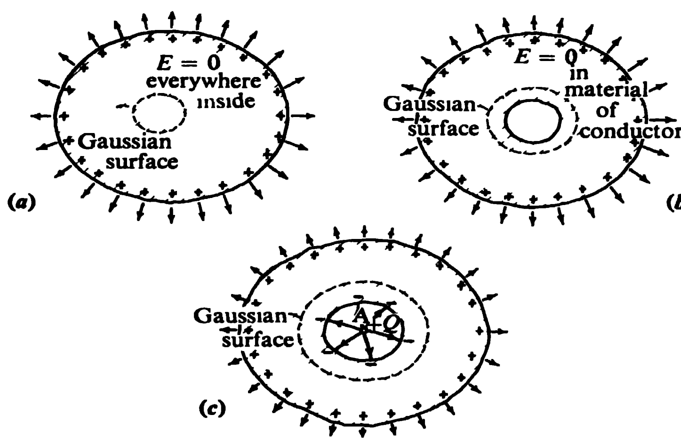


Fig. 4.2. Distribution of charge on conductors.

the surface while a hollow one (Fig. 4.2b) carries all its excess charge on the outer surface, unless there is a charge $+Q$ on a body A placed in the hollow (Fig. 4.2c), when a charge of $-Q$ must be

* At this stage some authors like to introduce the quantity $\mathbf{D}=\varepsilon_0\mathbf{E}$, the electric displacement *in vacuo*, when (4 2) becomes $\oint\!\!\oint \mathbf{D}\cdot d\mathbf{S}=\Sigma Q$ This gives expressions for **D** identical with those for **E** later in the chapter but with the ε_0 removed conversion to CGS e.s u or Gaussian units is then rendered difficult without a full consideration of the definition We shall postpone the introduction of **D** until a more natural point in chapter 13 Readers who wish may work in terms of **D** the MKSA unit is 1 C/m^2

induced on the inner surface. If the conductor in Fig. 4.2c is uncharged as a whole, a charge $+Q$ must be induced on the outer surface whatever the position of A. If A is a conductor and is allowed to touch the inside, the induced $-Q$ and the $+Q$ on A neutralize each other leaving the outer $+Q$ unaffected.

We can go further. The inner surface S (Fig. 4.3) of a charged hollow conductor is an equipotential. Suppose that there is no charge placed in the hollow but that nevertheless an electric field

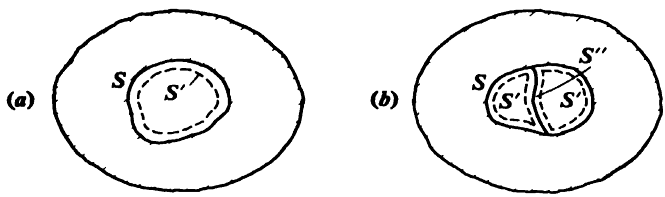


Fig. 4.3. Absence of electric field inside a hollow conductor.

exists so that there are points inside at potentials different from that of S. If all points just inside S are at a higher potential, then an equipotential surface like S' exists and lines of force between S and S' are all outwards: the flux of **E** over S' is not zero and a charge exists within it, which is self-contradictory. If all points just inside S are at a lower potential, a similar argument applies. If some are lower and some are higher, there exists a surface like S'' at the same potential as S dividing the hollow into two regions to which the same arguments apply since both are bounded by equipotentials. We can only conclude that **E** *is zero at all points within a hollow conductor not containing charge* and that the interior is an equipotential volume at the same potential as the conductor: all this is irrespective of happenings outside.

One last deduction. The fact that there is no *resultant* charge on

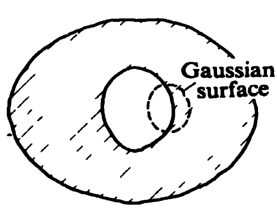


Fig. 4.4. Absence of charge on the inner surface of a hollow conductor.

the inner surface of the conductor in Fig. 4.2b does not guarantee the absence of equal and opposite charges. But *any* Gaussian surface chosen as in Fig. 4.4 to intersect part of the inner surface has $\mathbf{E}=0$ at all points and can contain no resultant charge.

Experimental Tests and Applications. That charge resides only on the outer surface of conductors was suggested in 1729 by Stephen Gray who showed that a solid and hollow cube of the same size produced the same electrical effects at distant points when charged in the same way: the absence of the inside made no difference

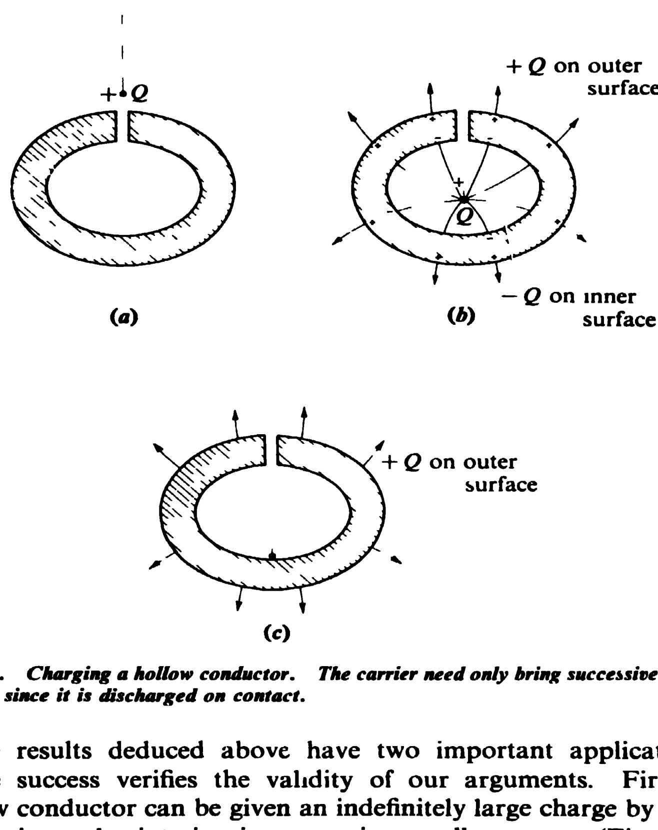


Fig. 4.5.* *Charging a hollow conductor. The carrier need only bring successive small charges since it is discharged on contact.

The results deduced above have two important applications whose success verifies the validity of our arguments. First, a hollow conductor can be given an indefinitely large charge by conveying it to the interior in successive small amounts (Fig. 4.5). Secondly, any apparatus placed inside a hollow conductor whose potential is fixed is unaffected by external electric fields: this is the

principle of *electrostatic shielding* in which the potential of the screen is usually fixed by a connection to the earth.

Any small apertures in the hollow conductor do not affect these results appreciably and the reader should now be able to explain the phenomena described in section 1.1 and known collectively as Faraday's ice-pail experiment.

4.2 Differential Equations for E and V

Gauss's theorem (4.2) and equation (3.17) represent general properties of electrostatic fields over a surface and round a path respectively. By applying the equations to elementary surfaces and paths and proceeding to the limit, relations can be obtained which apply *at a point.*

Poisson's and Laplace's Equations. Let us apply Gauss's theorem to a small element of space in cartesian co-ordinates occupied by charge of mean volume density ρ (Fig. 4.6). The elementary

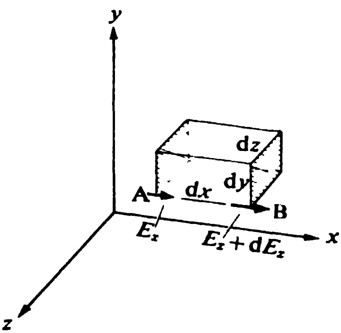


Fig. 4.6. Gauss's theorem applied to a volume element in cartesian co-ordinates.

region is a rectangular parallelepiped with faces parallel to the co-ordinate planes and with edges of length dx, dy and dz. Let the electric field strength at A have components E_x, E_y and E_z and consider the outward flux over the shaded surfaces (due only to the x-components of E because E_y and E_z are parallel to them). At B, the x-field is E_x+dE_x or $E_x+(\partial E_x/\partial x)\,dx$ and it should be clear that the x-field at any point on the right-hand shaded surface exceeds that at the corresponding point on the left-hand surface by $(\partial E_x/\partial x)\,dx$. The *resultant* outward flux from the two shaded surfaces is thus $(\partial E_x/\partial x)\,dx\,dy\,dz$.

Similar arguments applied to the other faces in pairs give a total outward flux of $(\partial E_x/\partial x+\partial E_y/\partial y+\partial E_z/\partial z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$, which must be equal to the enclosed charge, $\rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$, divided by ε_0. If the volume is allowed to shrink to A, then in the limit ρ will be the charge density at A and thus in general

$$\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z} = \frac{\rho}{\varepsilon_0} \tag{4.3}$$

at a point. This is a differential form of Gauss's theorem in cartesian co-ordinates and it can be expressed in terms of the potential V by using $E_x = -\partial V/\partial x$, etc., giving

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\varepsilon_0} \quad \text{(Poisson's equation)} \tag{4.4}$$

Where no charge exists

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2} = 0 \quad \text{(Laplace's equation)} \tag{4.5}$$

A Second Differential Equation for **E**. Now let us apply the general relation (3.17) to a path enclosing an elementary area in cartesian co-ordinates. The area in Fig. 4.7 is parallel to the xy-plane, has sides $\mathrm{d}y$ and $\mathrm{d}x$ as shown and is traversed from A in a

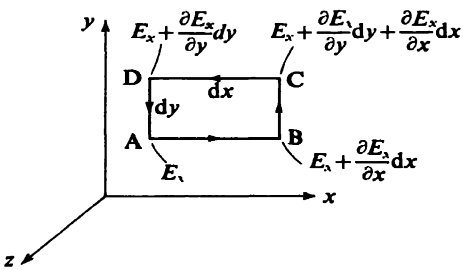


Fig. 4.7. Derivation of equation (4.6). The fields shown are only the x-components of **E**.

clockwise sense when looking along the positive z-axis (this is related to the general right-hand screw rule of appendix 7.1). The electric field at A has components E_x, E_y and E_z. The contributions to $\oint \mathbf{E}\cdot\mathrm{d}\mathbf{s}$ from AB and CD result only from the x-components whose values are shown in the figure at the corners. It is clear that

the field along CD always exceeds that at the same distance along AB by $(\partial E_x/\partial y)\,\mathrm{d}y$. Hence, because AB and CD are traversed in opposite directions, their total contribution to $\oint \mathbf{E}\cdot\mathrm{d}\mathbf{s}$ is $-(\partial E_x/\partial y)\,\mathrm{d}y\,\mathrm{d}x$. Similarly the contribution from BC and DA is $(\partial E_y/\partial x)\,\mathrm{d}x\,\mathrm{d}y$, and so from (3.17)

$$\partial E_y/\partial x - \partial E_x/\partial y = 0 \tag{4.6}$$

and similarly

$$\partial E_z/\partial y - \partial E_y/\partial z = 0$$

$$\partial E_x/\partial z - \partial E_z/\partial x = 0$$

These equations could have been more simply derived by using $E_y = -\partial V/\partial y$ and $E_x = -\partial V/\partial x$ and using the fact that $\partial^2 V/\partial x\,\partial y = \partial^2 V/\partial y\,\partial x$, but the method we have used is more useful later in cases when the right-hand sides are not zero.

4.3 Spheres of Charge

Spherical Shells. A conducting sphere of radius a carries all its charge Q on its surface and by its symmetry will constitute a spherical shell of uniform surface density of charge $\sigma = Q/4\pi a^2$. Suppose we want to find the electric field strength at a distance r from its centre. By the spherical symmetry of the problem if the field at P (Fig. 4.8) is E, then it must be radially outwards and have the same magnitude at all points on the surface of a sphere S of radius r concentric with the conductor. If S is chosen as a Gaussian surface, E is everywhere normal to it and has the same magnitude over it, so that the outward flux over S is $4\pi r^2 E$. By Gauss's theorem

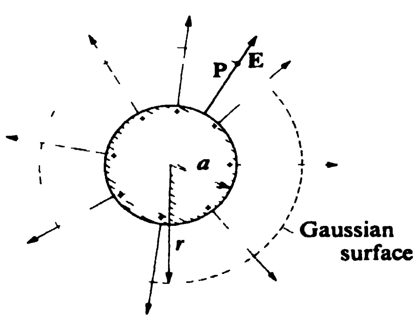


Fig. 4.8. Electric field due to a charged spherical conductor.

this is equal to Q/ε_0 and so

$$r \geqq a \qquad E = \frac{Q}{4\pi\varepsilon_0 r^2} \tag{4.7}$$

as if Q were concentrated at the centre. The potential will be

$$r \geqq a \qquad V = \frac{Q}{4\pi\varepsilon_0 r} \qquad \text{(zero at } \infty) \tag{4.8}$$

Inside the sphere $E = 0$ of course, and

$$r \leqq a \qquad V = \frac{Q}{4\pi\varepsilon_0 a} \qquad \text{(zero at } \infty) \tag{4.9}$$

If the surface density of charge is σ, the electric field *just outside* the surface is

$$E_{\text{surface}} = \frac{Q}{4\pi\varepsilon_0 a^2} = \frac{\sigma}{\varepsilon_0} \tag{4.10}$$

while the potential is

$$V_{\text{surface}} = \frac{Q}{4\pi\varepsilon_0 a} = aE_{\text{surface}} = \frac{a\sigma}{\varepsilon_0} \tag{4.11}$$

Equation (4.11) shows that, if we have a series of spheres of varying radii but all charged to the same potential, the surface densities of charge and the electric fields just outside the surfaces are inversely proportional to a, the radii of curvature. We might expect the same to happen on a single conductor whose surface (Fig. 4.9b) has various convex radii of curvature. The spheres of which the surfaces at points A, B and C are a part are all at the same potential, and we expect the surface density to be highest at A.* In general, therefore, charge should collect more densely

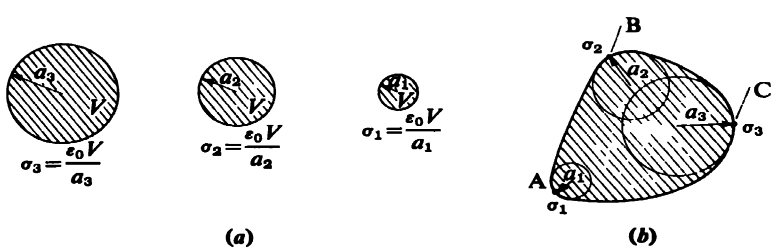


Fig. 4.9. ***(a) Spheres at the same potential have $\sigma \propto 1/a$; (b) a conductor with its surface at the same potential at all points has $\sigma \propto 1/a$ approx.***

* This is not intended as a proof, but merely as suggestive.

where the convex radius of curvature is smaller, and particularly at points. The electric field strength is correspondingly greater and at sharp corners may be large enough to cause the ionization of the air around them: ions of opposite sign are then attracted to the conductor and discharge it, while those of the same sign are repelled and cause a detectable wind. Discharge from points is known as *corona discharge* (section 12.6). While it can be usefully employed as in a lightning conductor or in the van de Graaff generator, it is to be avoided for obvious reasons in high voltage apparatus, whose external surfaces should all be free of corners and particles of dust.

Spherical Volumes of Charge. If we now take a spherical volume of radius a carrying a total charge Q distributed uniformly throughout it, the same argument as in the case of the shell will lead to (4.7) and (4.8) for the electric field and potential at points *outside*. At points *inside*, take a Gaussian surface S of radius r as in Fig. 4.10

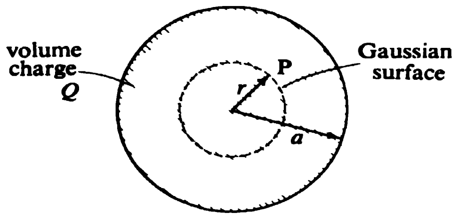


Fig. 4.10. The electric field inside a spherical volume of charge.

over which E is the same at all points: the outward flux is then still $4\pi r^2 E$, but the enclosed charge this time is

$$Q \times \frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi a^3} \quad \text{or} \quad Qr^3/a^3$$

and so

$$r \leqq a \qquad E = \frac{Qr}{4\pi\varepsilon_0 a^3} \tag{4.12}$$

To evaluate V, we use equation (3.12), taking a zero at infinity and remembering that the path for $r > a$ has E given by (4.7) while for $r < a$, E is given by (4.12). Thus, using x as the variable distance from the centre,

$$V = \int_\infty^a -\frac{Q\,\mathrm{d}x}{4\pi\varepsilon_0 x^2} + \int_a^r -\frac{Qx\,\mathrm{d}x}{4\pi\varepsilon_0 a^3}$$

i.e.

$$r \leqq a \qquad V = \frac{Q}{8\pi\varepsilon_0 a}\left(3 - \frac{r^2}{a^2}\right) \qquad \text{(zero at } \infty) \qquad (4.13)$$

Figure 4.11 compares the variations of E and V for a point charge, a spherical shell and a spherical volume of charge.

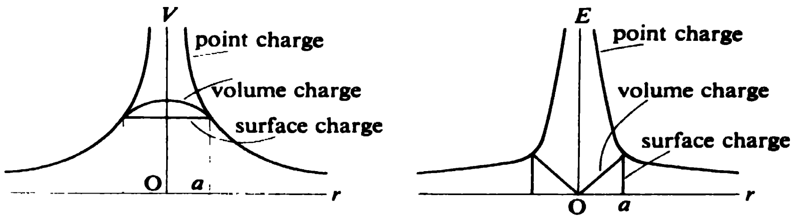


Fig. 4.11. Comparison of potential diagrams and variations of E with distance for a point charge and for surface and volume distributions of the same radius, all containing the same total charge.

The Nuclear Atom. As an application of the above results, we shall examine two models of the atom. The average volume occupied by an atom, calculated from Avogadro's number (problem 1.9), has a radius of the order of 10^{-8} cm. Knowing that negative electrons were constituents of all matter and that the atom as a whole was neutral, J. J. Thomson proposed (1907) an atomic model in which the positive charge was spread over a sphere of radius of about 10^{-8} cm with the electrons embedded in it. Rutherford, in projecting α-particles (known to be helium ions) at various substances, found that the deflections suffered were sometimes larger than could be accounted for by Thomson's atom. Accordingly, he suggested (1911) a model in which the positive charge was concentrated in a much smaller region of radius less than 10^{-12} cm, the *nucleus*, while the electrons surrounded it. This model predicted a law of scattering of α-particles which was confirmed in every detail by Geiger and Marsden (1913).

The path of an α-particle projected at a heavy positive nucleus can be calculated using equation (3.23)—$Q\mathbf{E} = m\mathbf{a}$—and, like all particle motion under an inverse square law of force, is a conic section, though in this case a hyperbola because of the repulsion. Rutherford's calculation can be followed in his original paper and we shall only look at the matter from one aspect: that of the inverse square law.

Figure 4.12 compares the potential energies of α-particles at various distances from the centre of a heavy atom, such as gold, for the Thomson and Rutherford models. The α-particles used by Rutherford had kinetic energies of about 8 MeV and any passing through or close to the centre of an atom would behave very differently in the two cases, as the figure shows (see appendix 4.2 for the use of potential diagrams in particle motion).

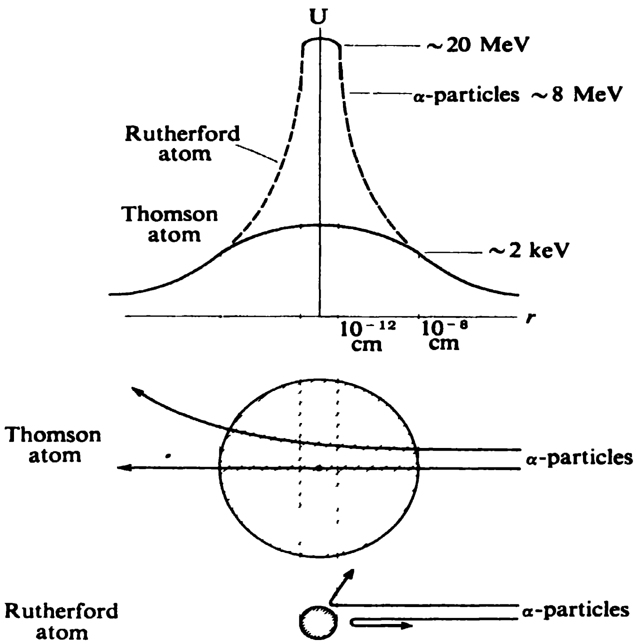


Fig. 4.12. Potential diagrams and α-particle trajectories for Thomson and Rutherford atoms. The numerical values are for a heavy atom such as gold, but the scales are grossly distorted to show the essential features clearly.

The Rutherford scattering law is predicted only by inverse square repulsion and, from experimental results confirming it, Chadwick (1920) showed that for heavy atoms the exponent in Coulomb's law does not differ from 2 by more than 0·03 down to distances of about 10^{-12} cm. Recent experiments with electrons and muons show that the law may be applied without detectable error down to 10^{-14} cm so that the assumption in section 2.5 is amply justified.

4.4 Cylinders of Charge

An infinite cylindrical conductor of radius a carries a charge λ per unit length distributed uniformly over its surface. To determine the electric field strength E at a point P a distance r from the axis, choose a Gaussian surface as in Fig. 4.13. By the cylindrical symmetry of the system, the field will be radially outwards and

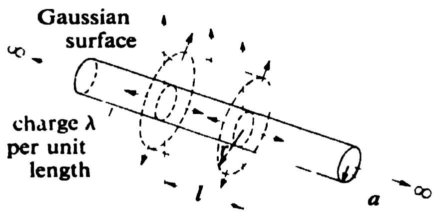


Fig. 4.13. The electric field due to an infinite charged conducting cylinder.

will have the same magnitude at all points on the curved surface of radius r. The outward flux of E is thus $2\pi rlE$, while the enclosed charge is λl, so that by Gauss's theorem

$$r \geqq a \qquad E = \lambda/2\pi\varepsilon_0 r \qquad (4.14)$$

which would not change if the charge were distributed along the axis. In practice we meet only finite cylinders but we expect (4.14) to be the more accurate the greater the ratio of length to diameter and the further P is from the ends (see problem 2.9).

The potential V cannot be evaluated with infinity as a zero because $\int_x^r -\mathbf{E}\cdot d\mathbf{s}$ becomes infinite. Any other point may be chosen and if we take the cylinder itself, using x again as the variable distance:

$$V = \int_a^r -\frac{\lambda\, dx}{2\pi\varepsilon_0 x}$$

$$= \frac{\lambda}{2\pi\varepsilon_0}\log_e (a/r) \quad \text{(zero at cylinder)}$$

or, in general,

$$V = C - \frac{\lambda}{2\pi\varepsilon_0}\log_e r \qquad (4.15)$$

where C is a constant determined by the zero of V.

If the cylindrical region has charge distributed throughout its volume, as might be the case with a beam of electrons for instance, the same results are obtained for the outside field and potential, but the inside field ($r<a$) is now not zero (see problem 4.8).

4.5 Planes of Charge

We require the electric field strength E at a point P outside an infinite conducting plane carrying a charge σ per unit area distributed uniformly over the surface. Symmetry demands that the field shall be normal to the plane of the surface and shall have the same value at all points the same distance from the plane as P. Thus, a Gaussian surface chosen as in Fig. 4.14 has outward flux

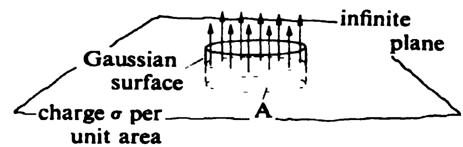


Fig. 4.14. The electric field due to an infinite charged conducting plane.

only across the upper face. This flux has magnitude EA, the enclosed charge is σA, and Gauss's theorem therefore gives

$$E = \sigma/\varepsilon_0 \tag{4.16}$$

a uniform field as predicted by equation (3.6). Once again, we do not meet infinite planes but we expect (4.16) to be sufficiently accurate for points near to a surface and some distance from the edges. The reader should check that for such points equations (4.7) and (4.14) both yield (4.16).

The potential cannot be evaluated with infinity as the zero but, since the field is uniform, equation (3.22) applies and

$$V = C - Ex = C - \sigma x/\varepsilon_0 \tag{4.17}$$

where C is a constant determined by the zero of V and x is the distance from the plane.

In this example, as in those of sections 4.3 and 4.4, we have assumed the conductors to be isolated in space and the lines of force to be going off to infinity, the implication being that these lines of force eventually end on equal negative charges. In practice, our formulae may be approximate not only because practical conduc-

tors are finite but also because the 'surroundings' are at a finite distance. The problem of other charges at finite distances is discussed in the next chapter.

4.6 The Electric Dipole

An electric dipole consists of two equal and opposite charges $\pm Q$ separated by a distance l: an *ideal* dipole is one in which l is negligible compared with distances of other charges but in which Q is correspondingly large so that Ql is still finite. We shall calculate fi.st the electric field due to a dipole, next the forces and couples on a dipole in an electric field and finally the interaction between two dipoles.

Electric Field due to an Ideal Dipole. To find the electric field strength at a point P with polar co-ordinates (r, θ) with respect to the centre of a dipole as in Fig. 4.15, we first calculate the potential.

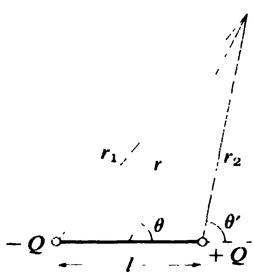


Fig. 4.15. Calculation of the potential at P due to an electric dipole.

From equation (3.15)

$$V_P = Q/4\pi\varepsilon_0 r_2 - Q/4\pi\varepsilon_0 r_1 \quad \text{(zero at } \infty)$$

$$= Q(r_1 - r_2)/4\pi\varepsilon_0 r_1 r_2 \tag{4.18}$$

But $r_1^2 = r_2^2 + l^2 + 2r_2 l \cos\theta'$ or $r_1^2 - r_2^2 = l(l + 2r_2\cos\theta')$ and, using this to substitute for $(r_1 - r_2)$ in (4.18), we obtain

$$V_P = \frac{Ql(l + 2r_2\cos\theta')}{4\pi\varepsilon_0 r_1 r_2 (r_1 + r_2)}$$

This is exact, but for an ideal dipole ($l \ll r$) $r_1 \to r$, $r_2 \to r$ and $\theta' \to \theta$, and in the limit

$$V_P = \frac{Ql\cos\theta}{4\pi\varepsilon_0 r^2} \tag{4.19}$$

A more elegant and more general method of deriving this expression is obtained by realizing that the potential at P is the increment in $Q/4\pi\varepsilon_0 r$ for a small displacement l of Q in the x-direction ($x = r\cos\theta$). Thus

$$V_P = -l\frac{\partial(Q/4\pi\varepsilon_0 r)}{\partial x} = -\frac{Ql}{4\pi\varepsilon_0}\frac{\partial(1/r)}{\partial x} = \frac{Ql}{4\pi\varepsilon_0 r^2}\,\partial r/\partial x \quad (4.20)$$

which is (4.19) because $\partial r/\partial x = x/r = \cos\theta$ (see problem 3.14). The negative sign occurs in (4.20) because it is Q which moves and not P.

The quantity Ql is the electric moment of the dipole or the *electric dipole moment*, denoted by p (unit C-m), so that

$$p = Ql \qquad \text{(Definition of } p\text{)} \quad (4.21)$$

and, from (4.19),

$$V_P = \frac{p\cos\theta}{4\pi\varepsilon_0 r^2} \quad (4.22)$$

Using (3.21), the components of **E** are

$$E_r = \frac{2p\cos\theta}{4\pi\varepsilon_0 r^3}; \qquad E_\theta = \frac{p\sin\theta}{4\pi\varepsilon_0 r^3} \quad (4.23)$$

giving a resultant field (Fig. 4.16) of magnitude $p(1+3\cos^2\theta)^{1/2}/4\pi\varepsilon_0 r^3$ at an angle ϕ to r such that $\tan\phi = \frac{1}{2}\tan\theta$.

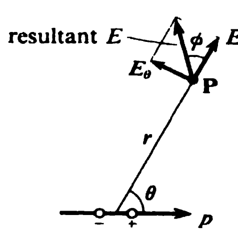


Fig. 4.16. The electric field due to a dipole.

If a positive *direction* from $-Q$ to $+Q$ is associated with p, then in Fig. 4.17, V_P due to p_1 and p_2 is $(p_1\cos\theta_1 + p_2\cos\theta_2)/4\pi\varepsilon_0 r^2$. If p_3 is the diagonal of the parallelogram, $p_3\cos\theta_3 = p_1\cos\theta_1 + p_2\cos\theta_2$ and $V_P = (p_3\cos\theta_3)/4\pi\varepsilon_0 r^2$. This last expression is also the V_P due to p_3 at θ_3 to r and hence p_3 is the resultant of p_1 and p_2:

dipole moments add like vectors and can be denoted by **p**. Equation (4.22) thus becomes

$$V_P = \frac{pr\cos\theta}{4\pi\varepsilon_0 r^3} = \frac{\mathbf{p}\cdot\mathbf{r}}{4\pi\varepsilon_0 r^3} \tag{4.24}$$

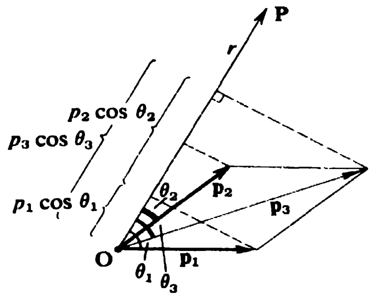


Fig. 4.17. Dipole moments add vectorially.

The electric fields of dipoles are shown in Figs. 4.18a and b, the former also including equipotential surfaces: note particularly that the plane bisecting the line joining the charges and perpendicular to it has $V=0$ (zero at ∞).

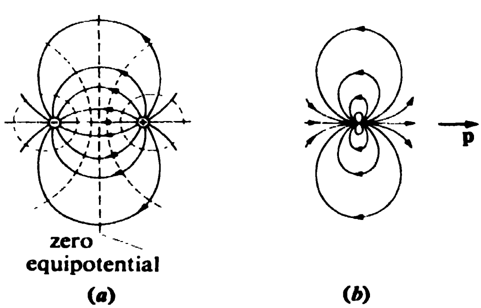


Fig. 4.18. (a) The electric field of an electric dipole; (b) The electric field of an ideal dipole.

Dipole in a Uniform Electric Field. In a uniform field **E**, the forces on the charges of the dipole are both $Q\mathbf{E}$ but are in opposite directions (Fig. 4.19) constituting a couple of moment $QEl \sin\theta$,

where θ is the angle between $\mathbf{p}$ and $\mathbf{E}$. Thus

$$T_\theta = pE \sin \theta \tag{4.25}$$

gives the couple or torque on a dipole in a uniform field, while the force

$$\mathbf{F} = 0$$

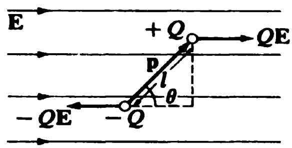


Fig. 4.19. The couple on a dipole in a uniform electric field.

Dipole in a Non-uniform Electric Field. The general case here is complex, so let us begin with a dipole lying along the x-direction in a field also in the x-direction but not uniform (Fig. 4.20a). If the field at $-Q$ has the magnitude E_x, then for a dipole of length $\mathrm{d}x$ the

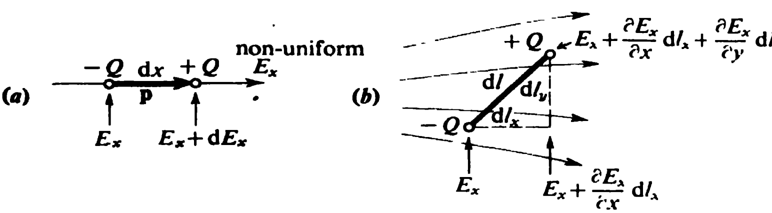


Fig. 4.20. Electric dipoles in non-uniform fields. In (b) the dipole is in the xy-plane: in general, a displacement $\mathrm{d}l_z$ will add a fourth term to the x-field at $+Q$.

field at $+Q$ will be $E_x+\mathrm{d}E_x$ or $E_x+(\mathrm{d}E_x/\mathrm{d}x)\,\mathrm{d}x$. The forces are now not equal and opposite but have a resultant $Q\,\mathrm{d}x(\mathrm{d}E_x/\mathrm{d}x)$. Since $Q\,\mathrm{d}x$ is the dipole moment p, the force is

$$F_x = p\frac{\mathrm{d}E_x}{\mathrm{d}x} \tag{4.26}$$

In general, the displacement of $+Q$ from $-Q$ will not lie along x but will be, say, $\mathrm{d}l$ with components $\mathrm{d}l_x$, $\mathrm{d}l_y$ and $\mathrm{d}l_z$. Moreover, the field will also have components E_x, E_y and E_z, all of which may vary. Consider the force in the x-direction, F_x, which must be due to

variations in E_x. If we use E_x to denote the x-component of $\mathbf{E}$ at $-Q$, then that at $+Q$ will be $E_x+(\partial E_x/\partial x)\,\mathrm{d}x+(\partial E_x/\partial y)\,\mathrm{d}y+(\partial E_x/\partial z)\,\mathrm{d}z$ and, although the QE_x's cancel, there is a resultant force

$$F_x = p_x\left(\frac{\partial E_x}{\partial x}\right)+p_y\left(\frac{\partial E_x}{\partial y}\right)+p_z\left(\frac{\partial E_x}{\partial z}\right) \tag{4.27}$$

because $Q\,\mathrm{d}l_x=p_x$, the x-component of $\mathbf{p}$, etc. Expressions for F_y and F_z are similar to (4.27).

The couple on an ideal dipole in a non-uniform field is still given by $T_\theta=pE\sin\theta$ because both forces are still $Q\mathbf{E}$, neglecting terms of a smaller order of magnitude.

Potential Energy of a Dipole in an Electric Field. A dipole of moment $\mathbf{p}$ makes an angle θ with an electric field $\mathbf{E}$ as in Fig. 4.21.

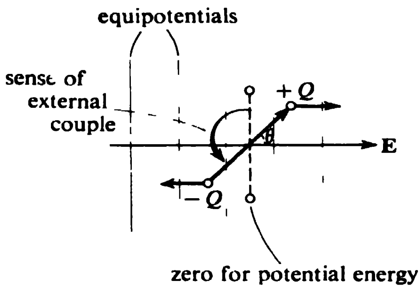


Fig. 4.21. Potential energy of a dipole in an electric field.

To calculate its potential energy a zero position must be chosen. The natural choice would seem to be when $\theta=0$, but it is more usual to take $\theta=90°$ for the following reason: since the equipotentials are at right angles to $\mathbf{E}$, the charges comprising the dipole lie on the same equipotential and any work done in establishing the dipole in the field is not counted (the work done in bringing up $+Q$ is equal and opposite to that done in bringing up $-Q$).

Starting then from $\theta=90°$, we now have to calculate the work done by external forces in rotating the dipole to its final position. By equation (4.25), the external couple necessary has a moment $pE\sin\alpha$ for a general angle α and will do work $pE\sin\alpha\,\mathrm{d}\alpha$ in rotating the dipole through a small angle $\mathrm{d}\alpha$. The work done in rotating it from 90° to θ is

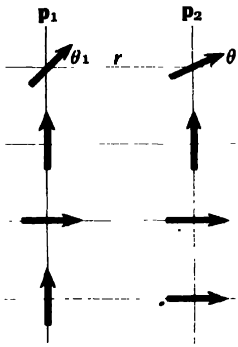
$$W' = \int_{\pi/2}^{\theta} pE\sin\alpha\,\mathrm{d}\alpha = [-pE\cos\alpha]_{\pi/2}^{\theta}$$

and the potential energy is thus

$$U = -pE\cos\theta = -\mathbf{p}\cdot\mathbf{E} \tag{4.28}$$

the negative sign occurring because work is done *on* the external couple.

Mutual Action between Dipoles. The general formulae for the forces and couples between two dipoles are complex and we shall limit ourselves to coplanar dipoles. A general method for finding these is left to a more advanced volume: here we shall use the vector properties of the dipole moment to resolve $\mathbf{p}_1$ and $\mathbf{p}_2$ of a general arrangement such as that of Fig. 4.22a along and perpendicular



	forces	couples	mutual potential energy
(a)			
(b)	$F_r = +\dfrac{3p_1p_2}{4\pi\varepsilon_0 r^4}$ (along r)	0	$\dfrac{p_1p_2}{4\pi\varepsilon_0 r^3}$
(c)	$F_r = -\dfrac{6p_1p_2}{4\pi\varepsilon_0 r^4}$ (along r)	0	$-\dfrac{2p_1p_2}{4\pi\varepsilon_0 r^3}$
(d)	$F_\theta = \pm\dfrac{3p_1p_2}{4\pi\varepsilon_0 r^4}$ (perpendicular to r)	$\dfrac{2p_1p_2}{4\pi\varepsilon_0 r^3}$ on p_1 (clockwise); $\dfrac{p_1p_2}{4\pi\varepsilon_0 r^3}$ on p_2 (clockwise)	0

***Fig. 4.22.** Mutual action between coplanar dipoles.*

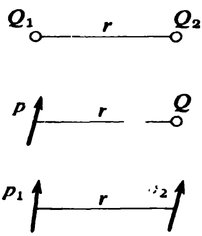
to r. The resultant interaction is then the sum of those between the components as in Fig. 4.22b, c and d. (The fourth case need not be considered separately since it is the same as Fig. 4.22d with the dipoles interchanged.)

The results quoted in Fig. 4.22 will not be derived completely here but sufficient indication of the methods will be given to enable the reader to do so. The couples and mutual potential energies are straightforward applications of (4.25) and (4.28) respectively. Particular notice should be taken of the couples in case (d) which are in the same sense on the two dipoles.

The forces, unlike the couples, must be equal and opposite on p_1

and p_2 so we need only consider one dipole, say p_2, in the (non-uniform) field of the other. Case (c) is the simplest, for here the moment and field are in the same direction and (4.26) will apply. Case (b) is left to problem 4.13. In case (d), the charges of p_2 are both in the 'broadside' fields of p_1, viz. $p_1/4\pi\varepsilon_0 r^3$, so the forces are perpendicular to r and differ by $p_2\ \partial(p_1/4\pi\varepsilon_0 r^3)/\partial r$ or $3p_1p_2/4\pi\varepsilon_0 r^4$. Note that in case (d) the resultant couple for the system as a whole is zero when both forces and mutual couples are included.

Finally, in Fig. 4.23, the interactions between charges and dipoles are summarized by listing the important parts of the expressions for forces and potential energy.



	force	potential energy
Q_1 — Q_2	$\propto Q_1Q_2/r^2$	$\propto Q_1Q_2/r$
p — Q	$\propto Qp/r^3$	$\propto Qp/r^2$
p_1 — p_2	$\propto p_1p_2/r^4$	$\propto p_1p_2/r^3$

Fig. 4.23. Interactions between charges and dipoles.

4.7 The Quadrupole and General Arrangements of Charge

At very large distances any collection of charges occupying a finite volume acts as if it were a point charge of $\Sigma\, Q$, where the sum is algebraic, so that the electric field falls off as the inverse square of the distance. If it happens that $\Sigma\, Q=0$ we now see that this does not necessarily mean that **E** will be zero because the system could be a dipole whose field falls off as the inverse cube of the distance. This $1/r^3$ field *may* be present even when $\Sigma\, Q\neq 0$ but it falls off so much more rapidly than the inverse square that it can often be neglected.

Suppose now that we go one stage further and consider systems for which not only does $\Sigma\, Q=0$ (a scalar sum), but $\Sigma\, \mathbf{p}=0$ as well (a vector sum). Typical arrangements for which this is true are shown in Fig. 4.24 and are known as *quadrupoles*: in particular, (b) is a *linear quadrupole*. The electric field should not be zero any more than it was with the dipole, but we shall show that the decrease with distance is even more rapid. The general quadrupole is dealt with

in more advanced texts: we shall calculate the potential due to a linear one only.

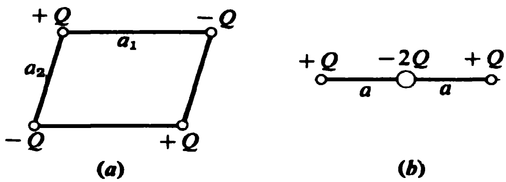


Fig. 4.24. (a) A general quadrupole; (b) a positive linear quadrupole.

Potential due to a Linear Quadrupole. Regard the quadrupole as derived from two equal but opposite dipoles of moment **p** placed in line as in Fig. 4.25 and separated by a distance a. Using the method

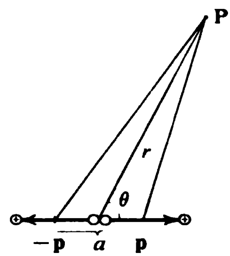


Fig. 4.25. Calculation of the potential due to a linear quadrupole treated as two opposite dipoles in line.

employed in equation (4.20), we want the increment in $(p\cos\theta)/4\pi\varepsilon_0 r^2$ for a small displacement of p in the direction of x. Thus

$$V_P = -a\frac{\partial(p\cos\theta/4\pi\varepsilon_0 r^2)}{\partial x} = \frac{-pa}{4\pi\varepsilon_0}\frac{\partial(\cos\theta/r^2)}{\partial x}$$

Using $\partial\theta/\partial x = -(\sin\theta)/r$ and $\partial r/\partial x = x/r = \cos\theta$ (see problem 3.14), this yields

$$V_P = pa(3\cos^2\theta - 1)/4\pi\varepsilon_0 r^3$$

The *quadrupole moment* q is defined as pa $(=Qa^2$ for the linear quadrupole) and thus

$$V_P = \frac{q}{4\pi\varepsilon_0 r^3}(3\cos^2\theta - 1) \tag{4.29}$$

E will therefore decrease as $1/r^4$.

General Arrangements of Charge. We can conceive of arrays for which $\Sigma\, Q=0$, $\Sigma\, p=0$ and $\Sigma\, q=0$, as in Fig. 4.26, forming *octopoles*, whose fields we should expect to fall off as $1/r^5$. It appears therefore as if any arrangement of charge will have an electric field with, in general, charge, dipole, quadrupole etc. components, but that in most cases the one which falls off least rapidly predominates so much that it is often the only one we need consider.

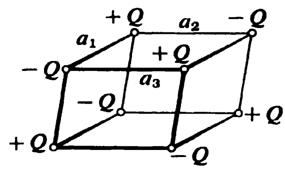


Fig. 4.26. A general octopole.

None of this affects the fields calculated in sections 4.3, 4.4 and 4.5 because for spherical symmetry etc. all the moments are zero. Any departure from spherical symmetry, as could occur in atomic nuclei for instance, will introduce small terms in **E** other than inverse square ones (see problem 4.16). Table 4.1 summarizes some properties of these terms, the last column showing that a charge may possess energy even when $\mathbf{E}=0$, that a dipole requires at least a uniform **E**, while a quadrupole requires a varying **E**.

Table 4.1

MULTIPOLES

	Potential V	*Field* $\mathbf{E}$	*Potential energy in electric field*
Charge, Q	$\propto Q/r$	$\propto Q/r^2$	$\propto QV$
Dipole, p	$\propto p/r^2$	$\propto p/r^3$	$\propto pE$ or $p\, \partial V/\partial x$
Quadrupole, q	$\propto q/r^3$	$\propto q/r^4$	$\propto q\, \partial E/\partial x$ or $q\, \partial^2 V/\partial x^2$, etc.

4.8 The Method of Images

In all the problems solved so far we have either been given a charge distribution or have been able to assume a uniform density because of a high degree of symmetry. In general, where the symmetry is lower or absent, the distribution on conducting surfaces must be found by other methods.

One such method is best illustrated by the simple example of Fig. 4.27a. Q is a point charge situated a perpendicular distance r from an infinite plane conducting surface which we shall take as a zero of potential. There is cylindrical symmetry about the line QO, but this only means that the surface density and field at P will

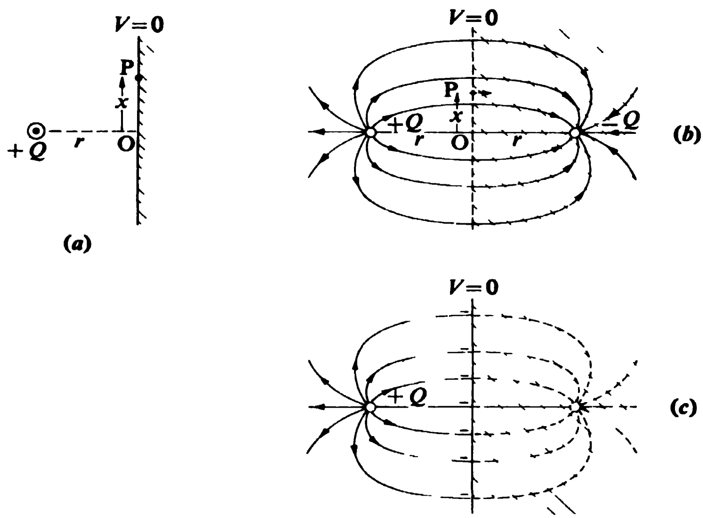


Fig. 4.27. (a) A point charge and an infinite conducting plane; (b) a point charge and an image $-Q$ in the plane; (c) lines of force (in full) due to a point charge and conducting plane: the dotted lines do not exist.

be the same as those at any other point on the surface a distance x from O.

Now look at the lines of force and equipotentials in Fig. 4.27b, a repetition of those in Fig. 4.18a. In the latter figure, the region to the left of the $V=0$ equipotential plane is a region of space containing a charge Q situated a perpendicular distance r from an infinite plane of zero potential, exactly the situation to the left of the conducting surface in Fig. 4.27a. The problem for this region is the same in the two cases and would be expected to give the same lines of force and equipotentials.

Thus to solve the given problem we can replace the plane by a charge $-Q$ known as the *image charge* since it is at the same point as the image of Q would be in a plane mirror through OP.

At P in Fig. 4.27b, the field is $E = 2Qr/4\pi\varepsilon_0(r^2+x^2)^{3/2}$ and, because $E = \sigma/\varepsilon_0$ outside any conducting surface,

$$\sigma_P = -2Qr/4\pi(r^2+x^2)^{3/2}$$

Moreover, the force between Q and the plane will be $Q^2/16\pi\varepsilon_0 r^2$ (notice that defining electric field strength as dF/dQ yields the contradiction referred to in section 3.1: an uncharged plane would give a field of $Q/8\pi\varepsilon_0 r^2$).

The general method of images thus involves replacing conducting surfaces by systems of charge which, together with the given charges, produce equipotential surfaces of the correct V identical with those of the original conductors. The fields and forces can then be calculated at any point and the surface densities of charge obtained from E as above.

The 'image system' is not necessarily the same as that in the equivalent optical system, as problem 4.20 shows.

4.9 Summary of Chapter 4

While this chapter has introduced no new experimental laws, Coulomb's law itself has been taken a stage further: Gauss's theorem expresses it in terms of the electric field strength $\mathbf{E}$

$$\oiint \mathbf{E}\cdot d\mathbf{S} = \Sigma\, Q/\varepsilon_0 \tag{4.2}$$

This, together with

$$\oint \mathbf{E}\cdot d\mathbf{s} = 0 \tag{3.17}$$

summarizes the properties of electrostatic fields. Equation (4.2) indicates that charges are *sources* of $\mathbf{E}$, i.e. that lines of force originate on charges as in Fig. 3.1a (negative sources are *sinks*). Equation (3.17) indicates that lines of $\mathbf{E}$ cannot close on themselves when they arise from static charges: this we could express by saying that there are no *vortices* of $\mathbf{E}$ in an electrostatic field (Fig. 7.1c shows a field with vortices).

The other fundamental results derived have been the differential forms

$$\partial E_x/\partial x + \partial E_y/\partial y + \partial E_z/\partial z = \rho/\varepsilon_0: \quad \partial E_y/\partial x - \partial E_x/\partial y = 0, \text{ etc.} \qquad \text{(4.3) and (4.6)}$$

and Poisson's and Laplace's equations

$$\nabla^2 V = -\rho/\varepsilon_0 \tag{4.4}$$

$$\nabla^2 V = 0 \tag{4.5}$$

where ∇^2 (read 'del squared'), known as the Laplacian operator, stands for $\partial/\partial x^2+\partial/\partial y^2+\partial/\partial z^2$ in cartesian co-ordinates

We have been rather glib about two matters One is our acceptance without question of the concept of an electric field within a volume of charge (equation (4 12)) why does the field not become infinite? This we shall examine in section 13 8 The other matter is the assumption on which the method of images and our arguments from symmetry rest and which we did not state explicitly that if a region of space is bounded by surfaces at known potentials and contains given charges, there is only one solution for the potential and electric field within the region This assumption is part of the *uniqueness theorem* whose proof must be left to a more advanced volume

Appendix 4.1 Flux of a Vector over a Surface

A small surface area dS can be treated as a vector whose direction is that of a line normal to it If dS is part of a larger area so that its two sides can be distinguished as *inside* and *outside* then let $\hat{\mathbf{n}}$ be a unit vector along the outward normal The *vector area* $\mathbf{dS}$ is defined as $\hat{\mathbf{n}}\, dS$ (Fig 4 28)

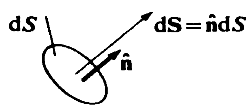


Fig 4 28 Surface area as a vector

Now let $\mathbf{A}$ be any vector quantity which has a value at all points on the surface of which $\mathbf{dS}$ is a part, and suppose $\mathbf{dS}$ to be small enough for $\mathbf{A}$ to be constant over it *The flux of* $\mathbf{A}$ *over* $\mathbf{dS}$ is defined as the product of the normal component of $\mathbf{A}$ and the area, i e $A\, dS \cos\theta$ or $\mathbf{A}\cdot\mathbf{dS}$ (Fig 4 29) It is thus a scalar quantity and will be negative if $\mathbf{A}$ crosses $\mathbf{dS}$ from outside to inside

If we require the flux of $\mathbf{A}$ over a large area S, we must add the elementary fluxes and obtain $\iint \mathbf{A}\cdot\mathbf{dS}$, where the integral is a

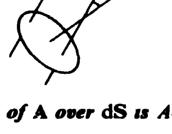


Fig. 4.29. The flux of $\mathbf{A}$ *over* $\mathbf{dS}$ *is* $A dS \cos\theta$ *or* $\mathbf{A}\cdot\mathbf{dS}$.

double or surface integral because an element of S is the product of two elements of distance. say dx and dy. The evaluation of surface integrals, like line integrals, should be looked up in mathematical texts if needed. We usually try to choose the surface in such a way that **A** is constant over some parts of it and zero over others, as in sections 4.3, 4.4 and 4.5.

In chapter 1, we could have regarded current as the flux of current density over a surface because, if **J** is the current density in magnitude and direction, the total current I over a surface S is

$$I = \iint_S \mathbf{J}\cdot d\mathbf{S} \tag{4.30}$$

so that, like any flux, current I may have a *sign* but is a scalar: current density is the vector (Fig. 4.30).

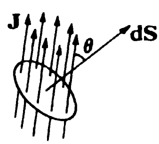


Fig. 4.30. Current as a flux of current density.

It should be made clear that flux does not necessarily mean the flow of anything even though the term derives from this idea.

Appendix 4.2 The Use of the Potential Diagram

In Fig. 4.31, the curve represents the potential energy of a particle at various distances from a centre of force. If the total mechanical energy is conserved, it can be represented by a horizontal line, W. Because the curve gives the potential energy, the kinetic energy is given by $W-U$ as at P in the figure.

A particle projected from infinity (where $U=0$) with a kinetic energy equal to W cannot approach nearer than R, at which point the kinetic energy, and hence the velocity, become zero. Another particle projected with an energy W_1 will suffer the same fate if it starts from a great distance: it will approach no nearer than r_1, but if captured by the centre of force it may oscillate between r_2 and r_3.

Because α-particles are repelled by a nucleus, the potential diagram of Fig. 4.12 was appropriate to show that the Rutherford atom would turn back some particles which would pass straight through the Thomson atom. On the other hand, an electron is

attracted by a nucleus and the appropriate potential diagram is now Fig. 4.32, in which an electron bound to the nucleus may move anywhere between r_1 and r_2 according to classical mechanics. These

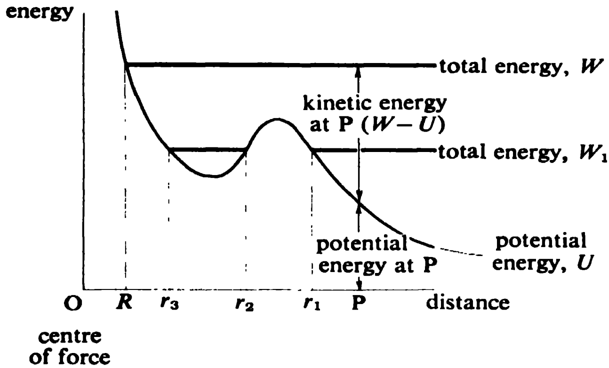


Fig. 4.31. The use of a potential diagram.

diagrams only give radial distances—motion with angular velocity about the centre of force is taking place simultaneously. (The quantum restriction on the energy of the atom means that W may

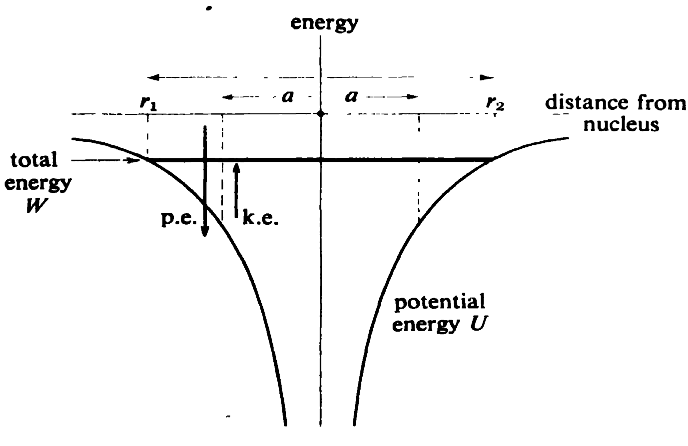


Fig. 4.32. Potential diagram for an electron bound to a nucleus. A Bohr circular orbit is one for which the potential energy is twice the kinetic energy in magnitude and opposite in sign.

only take on certain discrete values—the energy levels of the atom (see problems 2.11 and 2.12). In Fig. 4.32, the Bohr radius for a hydrogen atom is shown.)

References

The scattering of α-particles by the nuclear atom is probably best described in Rutherford's original paper (1911), but adequate accounts are to be found in many texts on nuclear physics.

PROBLEMS

SECTION 4.1

*4.1 The absence of an electric field inside a charged hollow conductor can be justified by a method different from that given in section 4.1. Consider the possible existence of electric lines of force inside such a conductor which (*a*) begin and end on the conductor, (*b*) begin on the conductor and end in space, (*c*) close on themselves or (*d*) do something other than (*a*), (*b*) or (*c*).

SECTION 4.2

4.2 Show that an electric field *in vacuo* whose lines of force are straight lines all parallel to the y-axis cannot vary in the x-direction.

*4.3 Apply Gauss's theorem to an element of volume in cylindrical polar co-ordinates and hence show that Poisson's equation in these co-ordinates is

$$\frac{1}{r}\,\partial V/\partial r+\partial^2 V/\partial r^2+\frac{1}{r^2}\,\partial^2 V/\partial\theta^2+\partial^2 V/\partial z^2 = -\rho/\varepsilon_0$$

Hence show that if cylindrical symmetry is complete (no variation with θ) and all cross-sections perpendicular to the z-axis are identical (no variation with z), V must be of the form $k\log_e r+C$, where k and C are constants.

SECTION 4.3

4.4 A conducting sphere 18 cm in diameter is charged to 9 kV above earth potential and isolated. Find the charge on the sphere, the electric field strength just outside it, and the field and potential 18 cm from its centre.

4.5 If air ionizes at a potential gradient of 30 kV/cm, what is the greatest charge which can be carried by a sphere of 18 cm diameter without the occurrence of corona?

4.6 If some apparatus is to be operated at 90 kV, what is the smallest external radius of curvature which should be permitted? (Dielectric strength of air = 30 kV/cm.) If smaller radii are unavoidable, what is a possible solution to the difficulty?

4.7 A 5 MeV α-particle makes a head-on collision with a gold atom. What is the distance of closest approach? Draw a potential diagram to scale to illustrate the collision. (Charge on α-particle $= 2e$; charge on gold nucleus $= 79e$.)

SECTION 4.4

4.8 Assuming that a beam of electrons can be treated as a charge distributed uniformly over a cylindrical region of radius a and of infinite length, find the electric field strength at a distance r from the axis for $r>a$ and $r<a$, taking the line density of charge as λ. If there are n electrons per unit volume in the beam, show that the force on an electron at a distance r from the axis is $ne^2r/2\varepsilon_0$ radially outwards, e being the electronic charge.

SECTION 4.5

4.9 A sheet of charge, surface density σ, gives an electric field of $\sigma/2\varepsilon_0$ on both sides. If, however, a thin sheet of metal is charged with σ per unit area, equation (4.16) seems to show that the field outside will be σ/ε_0. Resolve the paradox.

4.10 An infinite plane sheet of charge gives an electric field $\sigma/2\varepsilon_0$ at a point P a distance a from it. Show that half the field is contributed by charge whose distance from P is less than $2a$; and that in general all but $f\%$ of the field is contributed by charge whose distance from P is less than $100a/f$.

*4.11 A conductor carries its charge in a thin sheet on its surface, and the electric field just inside and outside the sheet should be $\sigma/2\varepsilon_0$ as in Fig. 3.4. How do you reconcile this with the zero field which in fact occurs in conducting material and the field of σ/ε_0 which occurs just outside?

SECTION 4.6

4.12 Show that a dipole of moment p situated at the origin and along the x-axis of rectangular cartesian co-ordinates produces components of electric field strength at the point (x, y) given by

$$E_x = p(2x^2-y^2)/4\pi\varepsilon_0 r^5; \qquad E_y = 3pxy/4\pi\varepsilon_0 r^5$$

where $r=(x^2+y^2)^{1/2}$.

*4.13 Use the result of problem 4.12 to solve case (b) of Fig. 4.20. (Place the second dipole along the y-axis a distance R from and parallel to the first. Use the method adopted for deriving equation (4.27), evaluating the force components at $x=0$, $y=R$.)

4.14 Show that, if the electric field strength at a point (r, θ) due to a dipole at the origin along the $\theta=0$ axis is resolved into two components, one along r (E_r) and the other parallel to the dipole ($E_{\parallel}$),

$$E_r = (3p\cos\theta)/4\pi\varepsilon_0 r^3; \qquad E_{\parallel} = -p/4\pi\varepsilon_0 r^3$$

4.15 Show that the equations of the lines of force of a dipole are $(\sin^2\theta)/r = \text{constant}$, and of equipotentials are $(\cos\theta)/r^2 = \text{constant}$.

SECTION 4.7

*4.16 Show that point charges $+Q$ situated at $(\pm a, 0)$ in the xy-plane give an electric field at large distances the same as that of a charge and a positive quadrupole at the origin. Hence argue that an atomic nucleus in

the shape of a spheroid elongated in the x-direction (prolate) would possess a positive quadrupole moment, that one contracted in the x-direction (oblate) would possess a negative quadrupole moment, and that neither would possess an electric dipole moment.

SECTION 4.8

4.17 Two very large conducting planes intersect at right angles and a point charge Q is situated in the corner a perpendicular distance a from both. Find the set of image charges making *both* planes zero equipotentials and find the force on Q.

4.18 Show, by integration over the surface, that the charge induced on the plane in Fig. 4.26 is $-Q$.

4.19 A simple pendulum has a mass m at the end of its string of length l. The mass possesses a charge Q and swings a distance a above a large horizontal conducting plane at zero potential. Find the period of oscillations whose amplitude is much smaller than a or l, neglecting all damping.

*4.20 A point charge Q is a distance d from the centre O of an earthed conducting sphere of radius a. Show that a charge $-Qa/d$ at a distance a^2/d from O along OQ is a satisfactory image charge and find the ratio of the maximum to the minimum surface density of charge induced on the sphere.

APPENDIX 4.2

4.21 The ionization potential of atomic hydrogen is 13·6 V. Find the radius of the electronic orbit on the Bohr model. (Use problem 2.11.)

4.22 Find the potential energy for a particle of mass m which can move only along the x-axis and is subject to a restoring force $-\mu x$. Plot a potential energy diagram and show from it that, if the amplitude of simple harmonic oscillations is a, the velocity at any point is $[\mu(a^2 - x^2)/m]^{1/2}$ and the total energy $\frac{1}{2}\mu a^2$.

CHAPTER 5

CAPACITANCE AND ELECTRIC ENERGY

The methods developed in the last two chapters can now be used to investigate more practical situations than those encountered so far. The important concept of capacitance, introduced as a quantity relating charge and potential difference and finding its embodiment in condensers, leads to a consideration of electric energy, of forces and couples between charged conductors and of electrostatic instruments.

5.1 Capacitance of a Conductor

When a finite conductor remote from other bodies is given a charge Q, its potential V (zero at infinity) increases in magnitude and the ratio is said to be the *capacitance of the conductor*, C, so that

$$C = Q/V \qquad \text{(Definition of } C\text{)} \qquad (5.1)$$

An older name for C is *capacity*, used because it gives the amount of charge the conductor holds at a certain potential. The definition enables C to be calculated for some particular conductors: for instance, equation (4.9) shows that for a conducting sphere of radius a

$$C = 4\pi\varepsilon_0 a \qquad (5.2)$$

though for other shapes an accurate calculation is not easy.

The MKSA unit of capacitance is the C/V or *farad*, symbol F, although the sub-multiples 1 μF (10^{-6} F) and 1 pF (10^{-12} F) are more convenient in practice.

The Proportionality of Q and V. Equation (5.2) shows that Q and V are proportional for a given conducting sphere: we shall now deduce this proportionality for a conductor of any shape or size and thus prove the constancy of C, the capacitance. There are two steps.

First, we show that, if the potential V of a conductor such as that in Fig. 5.1 is specified, there is only one distribution of charge over

the surface which gives this potential. For suppose that the surface density at any point on the surface a distance r from P is σ (not necessarily constant): the potential at P, and therefore of the whole conductor, is

$$V = \iint_S \frac{\sigma \, dS}{4\pi\varepsilon_0 r} \qquad \text{(zero at } \infty) \qquad (5.3)$$

by equation (3.16), where the integration extends over the surface of the conductor. Suppose also that another distribution σ' gives the same potential so that

$$V = \iint_S \frac{\sigma' \, dS}{4\pi\varepsilon_0 r} \qquad \text{(zero at } \infty) \qquad (5.4)$$

Subtracting (5.4) from (5.3) yields

$$\iint_S \frac{(\sigma - \sigma') \, dS}{4\pi\varepsilon_0 r} = 0 \qquad \text{(zero at } \infty) \qquad (5.5)$$

The left-hand side of this equation is an expression giving the potential the conductor would have if it possessed a distribution of

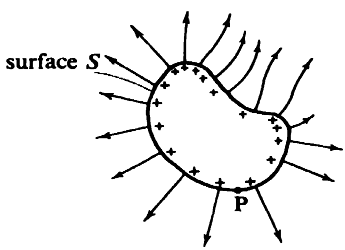


***Fig. 5.1.** A single charged conductor remote from other bodies. Lines of force end on surroundings effectively at infinity.*

charge of surface density $(\sigma - \sigma')$, and (5.5) tells us that this potential is zero. The space between the conductor and infinity is thus a volume containing no charges and bounded by an equipotential surface (in this case $V = 0$), and the argument of section 4.1 showed that in such a volume $\mathbf{E} = 0$ everywhere. By equation (4.16) the surface density of charge must then be zero, so that $\sigma = \sigma'$. Hence, if V is specified, σ is unique.

The second step is to consider our conductor first with a surface

density σ_1 producing a potential V_1 and then with a surface density σ_2 and a potential V_2, so that

$$V_1 = \iint_S \sigma_1 \, dS/4\pi\varepsilon_0 r \quad \text{and} \quad V_2 = \iint_S \sigma_2 \, dS/4\pi\varepsilon_0 r.$$

If the ratio of V_2 to V_1 is n then

$$V_2 = nV_1 = \iint n\sigma_1 \, dS/4\pi\varepsilon_0 r,$$

showing that V_2 is produced both by distributions $n\sigma_1$ and σ_2. By the first step above, $\sigma_2 = n\sigma_1$ and so $Q_2 = \iint \sigma_2 \, dS = \iint n\sigma_1 \, dS = nQ_1$. Hence Q and V are proportional and C is constant for a conductor of given shape and size.

5.2 Capacitance of Condensers

A conductor remote from all others as in Fig. 5.1 will generate lines of force from its charge Q which will end on charges of total amount $-Q$ effectively at infinity, the latter charges not affecting C because the zero for V is itself infinity. Most conductors however are influenced by other bodies, both conductors and insulators, and any charge on such bodies of the same sign as Q in (5.1) will raise V and thus diminish C. To increase the capacitance of a conductor therefore requires the proximity of charges of opposite sign.

Leaving the influence of insulators until chapter 13, we shall begin by considering the important situation illustrated in Fig. 5.2 in

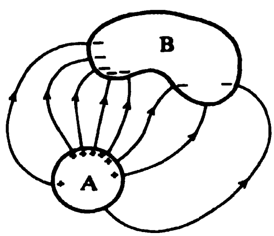


Fig. 5.2. An ideal condenser. All the lines from A *end on* B.

which only one other conductor B is near enough to influence the original conductor A and all the lines of force from A end on B. The charges on the conductors must be equal and opposite as if A were charged from B. Such an arrangement is known as a *condenser* or *capacitor* whose important property is now to hold the charge transferred at a low potential *difference*.

The capacitance of a condenser is defined by

$$C = \frac{Q}{V_A - V_B} \quad \text{(Definition of } C\text{)} \quad (5.6)$$

where V_A and V_B are the potentials of the originally uncharged conductors produced by the transference of charge Q from one to the other. Definition (5.1) is now seen to be a special case of (5.6) in which B surrounds A and is at a large distance away. For instance, two concentric spherical conducting shells of radii a and b as in Fig. 5.3 develop a potential difference between them of $Q/4\pi\varepsilon_0 a - Q/4\pi\varepsilon_0 b$ when Q is transferred from the outer to the inner. By (5.6)

$$C = \frac{4\pi\varepsilon_0 ab}{b-a} \quad (5.7)$$

which tends to (5.2) as b tends to infinity. Note particularly that the surroundings in section 5.1 cannot be ignored since there must be a destination for the lines of force.

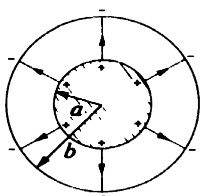


Fig. 5.3. A spherical condenser.

The constancy of C in the absence of insulators now follows by an extension of the argument at the end of the previous section. The potential V_A due to Q on A and $-Q$ on B can be shown to be proportional to Q by extending the integrals to cover both surfaces. V_B is similarly proportional to Q. The constancy of C thus rests very much on that part of Coulomb's law denoted by $F \propto Q_1 Q_2$ since it is this which yields the numerator of the expression $\iint \sigma \, dS/4\pi\varepsilon_0 r$ for potential. An experiment in which Q is measured by a ballistic galvanometer (section 16.6) and the potential difference by a potentiometer (section 6.8) is easily performed in the laboratory and confirms more accurately the law of section 2.2.

Condenser Geometry. An ideal condenser is best realized by

enclosing one conductor within the other as in the spherical condenser of Fig. 5.3, but this arrangement is inconvenient because of the inaccessibility of the inner conductor. The shapes in common use are the parallel plate and the cylindrical (Figs. 5.4 and 5.5) and

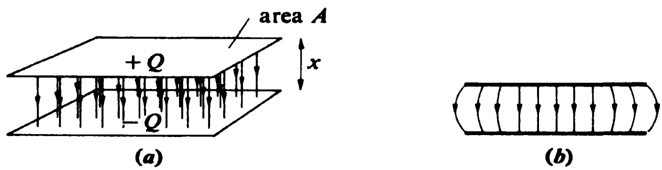


Fig. 5.4. A parallel-plate condenser (a) ideal; (b) showing edge effects.

because neither of these can ensure that no lines of force leak to the surroundings we shall only be able to deal with ideal conditions for the moment and assume no leakage. We shall also assume that the electric fields within the condensers are the same as if they were of infinite extent, which means that we neglect edge effects (sometimes known as fringing). The expressions obtained will be the more accurate as the distance between the conductors decreases in comparison with other linear dimensions and will be highly accurate if guard-rings are used (see section 5.4).

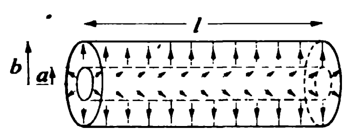


Fig. 5.5. A cylindrical condenser.

In Fig. 5.4, the transference of charge Q from one plate to the other produces surface densities of charge Q/A and hence an electric field of strength $Q/\varepsilon_0 A$ (equation (4.16)). The potential difference is thus $Qx/\varepsilon_0 A$ and hence

$$C = \varepsilon_0 A/x \tag{5.8}$$

for a parallel-plate condenser.

In Fig. 5.5, the transference of Q from the outer to the inner cylinder produces a charge per unit length of Q/l and thus a field of $Q/2\pi\varepsilon_0 rl$ at any distance r from the axis, the outer cylinder produc-

ing no field inside itself. The potential difference is thus $\int_b^a -Q\,\mathrm{d}r/2\pi\varepsilon_0 rl$ or $(Q \log_e b/a)/2\pi\varepsilon_0 l$. Hence

$$C = 2\pi\varepsilon_0 l/\log_e (b/a) \tag{5.9}$$

for a cylindrical condenser.

Note that because $C = Q/V$ a capacitance must always be ε_0 multiplied by an expression with the dimensions of a length. All such expressions can be converted to CGS e.s.u. by the substitution $\varepsilon_0 = 1/4\pi$. The dimensions of ε_0 are clearly those of capacitance per unit length, so that its value in MKSA units should be expressed as $10^{-9}/36\pi$ F/m.

Distributed Capacitance. In a condenser the capacitance is deliberately localized within a relatively small volume, but in extended conductors, such as cables or transmission lines used to convey electric currents over large distances, the capacitance is distributed continuously and is an important factor in any electric phenomena which occur.

Transmission lines, treated in detail in chapter 10, consist most often of two conductors arranged either as a coaxial cable (Fig. 5.6)

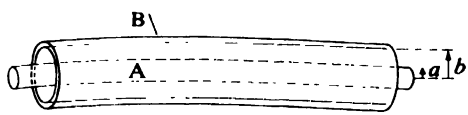


Fig. 5.6. A coaxial cable.

or as a twin cable (Fig. 5.7). The coaxial cable forms an extended cylindrical condenser and we expect (5.9) to apply quite accurately here in the form

$$C = 2\pi\varepsilon_0/\log_e (b/a) \text{ per unit length} \tag{5.10}$$

The capacitance of twin cable can be found most easily if the diameter of the wires is small compared with their distance apart

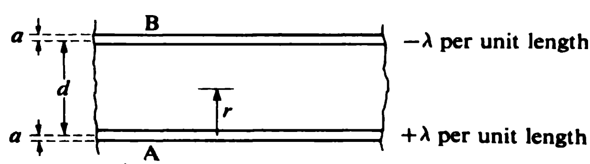


Fig. 5.7. A twin cable.

because the distribution of charge over the conductors can be neglected. If a charge λ per unit length is transferred from one to the other, the electric field strength at a point a distance r from one wire is $\lambda/2\pi\varepsilon_0 r - \lambda/2\pi\varepsilon_0(d-r)$. The potential difference

$$V_A - V_B = \frac{\lambda}{2\pi\varepsilon_0}\int_{d-a}^{a}\left[-\frac{1}{r}+\frac{1}{d-r}\right]\mathrm{d}r$$

$$= \frac{\lambda}{\pi\varepsilon_0}\log_e\frac{(d-a)}{a}$$

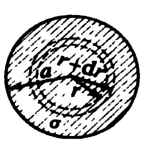
Hence $$C = \frac{\pi\varepsilon_0}{\log_e\frac{(d-a)}{a}} \approx \frac{\pi\varepsilon_0}{\log_e(d/a)} \text{ per unit length} \quad (5.11)$$

for twin cable in which $d \gg a$.

A single horizontal telegraph or aerial wire at a height h above the earth's surface would, if charged, have an electric image at a depth h. The capacitance per unit length between such a wire of radius a and the earth will thus be $\pi\varepsilon_0/\log_e(2h/a)$ from (5.11).

Estimation of Capacitance. Very often a rough estimate of capacitance is sufficient for practical purposes and the formulae already developed are then good enough approximations. They all apply, however, to cases in which the charge is at least approximately uniformly distributed and so, since the ability to estimate capacitance is a useful facility, we look at an example to which they do not apply at all.

A thin conducting disc of radius a would have a greatest charge density at the edges where the curvature is large. The distribution will in fact be intermediate between a uniform one and one in which *all* the charge is concentrated round the rim. A very simple calculation of the potential at the centre of the disc (and therefore of the disc as a whole) for these two extremes can be carried out.



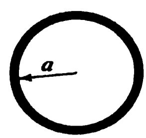
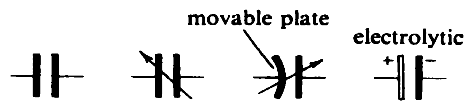


Fig. 5.8. Estimation of the capacitance of a disc. The true distribution of charge lies between the uniform distribution (a) and one in which it lies completely round the rim (b).

For a uniform charge density σ, a typical annulus of radius r (Fig. 5.8a) carries a charge $\sigma 2\pi r\,dr$ which produces a potential $\sigma\,dr/2\varepsilon_0$ at the centre. The potential due to the whole disc is thus $\sigma a/2\varepsilon_0$ and, since the charge is $\pi a^2\sigma$, the capacitance is $2\pi\varepsilon_0 a$. If all the charge is round the edge (Fig. 5.8b) the capacitance is similarly $4\pi\varepsilon_0 a$ so that the true value should lie between these two and we have at least an order of magnitude. (Accurate calculation gives $8\varepsilon_0 a$.)

5.3 Combinations of Condensers

We now consider more general cases in which several ideal condensers are connected by conducting wires into a network such as that of Fig. 5.9b. (Diagrammatically a condenser is represented by a pair of parallel lines, crossed by an arrow if variable: Fig. 5.9a.)

(a)


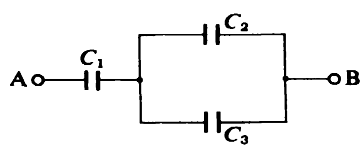
(b)


Fig. 5.9. (a) Condenser symbols; (b) a condenser network.

The *equivalent capacitance* between two points such as A and B is again defined by $Q/(V_A - V_B)$ where $V_A - V_B$ is the potential difference produced by the transference of charge Q from B to A. This is a slight generalization of definition (5.6) and means that we are asking for the capacitance of that one condenser which, if connected between A and B, would produce the same effect as the network. To obtain the equivalent capacitance we use two principles already established: the conservation of charge and the path-independence of potential difference.

Condensers in Series. A set of condensers as in Fig. 5.10 are in series and a charge Q transferred from one end to the other produces charges $\pm Q$ on the plates as shown. The potential differences are therefore $V_1 = Q/C_1$, $V_2 = Q/C_2$ and $V_3 = Q/C_3$. The

total V across the ends of the combination is the sum of these, and the equivalent capacitance C is thus given by

$$1/C = V/Q = (V_1+V_2+V_3)/Q = 1/C_1+1/C_2+1/C_3.$$

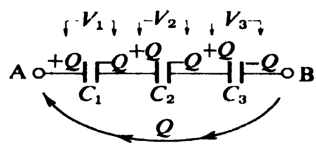


Fig. 5.10. Condensers in series.

In general therefore

$$\frac{1}{C} = \sum_i \frac{1}{C_i} \tag{5.12}$$

for condensers in series. Note particularly that these condensers act as a *potential divider*, with potential differences each inversely proportional to their capacitance.

Condensers in Parallel. In Fig. 5.11 it is the charge Q which

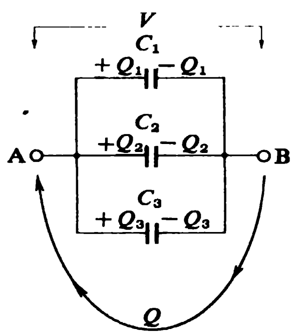


Fig. 5.11. Condensers in parallel.

divides into Q_1, Q_2 and Q_3 while all condensers have the same potential difference. The equivalent capacitance is

$$C = Q/V = (Q_1+Q_2+Q_3)/V = C_1+C_2+C_3$$

or in general

$$C = \sum_i C_i \tag{5.13}$$

for condensers in parallel.

General Networks. Many networks can be treated as a collection of series and parallel combinations and the equivalent capacitance found by using (5.12) and (5.13): the network of Fig. 5.9b, for instance, is equivalent to $C_1(C_2+C_3)/(C_1+C_2+C_3)$. But a network like that of Fig. 5.12 cannot be treated in this way—no two condensers are in series (common Q) or in parallel (common V)—so a general method based on the two principles mentioned above must be used.

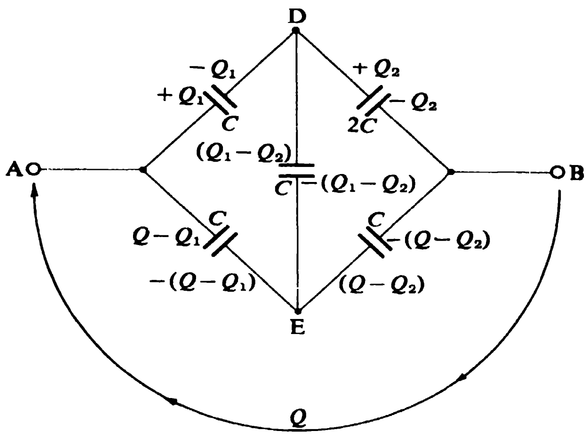


Fig. 5.12. A network which cannot be broken into series and parallel combinations.

Taking the network of Fig. 5.12 as an example, we again transfer Q from B to A. Let Q_1 and Q_2 be the charges on the condensers shown and apply the conservation of charge to the junctions A, B, D and E, giving the charges shown also in the figure. To find Q_1 and Q_2 apply the path-independence of V in the form: potential difference round a closed path is zero. This gives

$$Q_1/C-(Q-Q_1)/C+(Q_1-Q_2)/C = 0 \quad \text{round DAE}$$

$$Q_2/2C-(Q_1-Q_2)/C-(Q-Q_2)/C = 0 \quad \text{round BDE}$$

Solved for Q_1 and Q_2, these equations give $Q_1=7Q/13$ and $Q_2=8Q/13$. The potential difference between A and B via D is thus $11Q/13C$ and the equivalent capacitance is $13C/11$.

A further method for reducing networks is given in problem 5.7.

5.4 Condensers in Practice

The space between the plates of a real condenser is rarely a vacuum as we have assumed but is usually filled with an insulator even if only air. These insulators or *dielectrics* have several important effects which are detailed in chapter 13 but must be outlined here since they affect the design of condensers.

First, dielectrics increase the capacitance. If C_m is the capacitance of a condenser with a homogeneous insulating material filling the whole of the region in which an electric field exists and if C_0 is the capacitance of the same condenser *in vacuo* then

$$C_m/C_0 = \varepsilon_r \qquad \text{(Definition of } \varepsilon_r\text{)} \qquad (5.14)$$

defines the *relative permittivity* or *dielectric constant* of the material. Thus for a parallel-plate condenser whose plates are separated by an insulator of relative permittivity ε_r the capacitance is $\varepsilon_r\varepsilon_0 A/x$. (For some materials ε_r is not a constant and C_m, unlike C_0, is not dependent only on the geometry.)

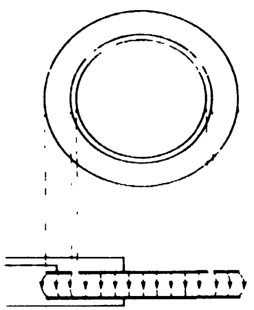
A second effect produced by an insulator is a limitation on the potential difference which can be applied across the plates because at some potential gradient known as the *dielectric strength* of the material it breaks down and conducts. For the moment the following orders of magnitude are sufficient: common gases at S.T.P. have ε_r between 1 and 1·001 and dielectric strengths of about 3 kV/mm, while common solid insulators have ε_r between 2 and 10 and dielectric strengths from a few kV/mm to several hundred kV/mm (see table 13.1).

Because ε_r may vary with applied potential difference, with pressure, temperature, humidity, frequency and with age, the characteristics of real condensers are much more variable than those of ideal ones and are determined almost entirely by the dielectric. Briefly, the types of condenser manufactured are designed to give the required capacitance in the smallest volume subject to the following: (*a*) whether the need is a fixed or continuously variable capacitance, (*b*) the largest potential difference which must be safely applied across the plates (the working voltage), (*c*) the required stability under varying physical conditions and with time, (*d*) the losses in energy (dielectric loss) and maximum current which can be permitted, no insulator being perfect. For small distances between the plates, capacitances are always given effectively by the parallel-plate formula $\varepsilon_r\varepsilon_0 A/x$, so that large A (obtained by stacking plates),

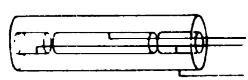
small x (obtained by using thin films of sufficient dielectric strength) and large ε_r all contribute to compactness.

Air and Vacuum Condensers. Fixed air condensers can in principle be used as primary standards since their capacitance can be calculated from dimensions if ε_0 is known, but there are good reasons for using inductances instead (section 16.4). Such condensers are therefore calibrated in farads by methods of measurement dealt with in chapter 10 and are then used as secondary standards.

Condensers of known dimensions can be used to measure ε_0 (section 16.7) but edge effects must then be avoided by using guard-rings or guard-cylinders (Fig. 5.13) maintained accurately in position and at the same potential as the conductor associated with



(a)



(b)

Fig. 5.13. Guard-ring and guard-cylinder condensers.

them, but not connected into the circuit. The only disturbing factors are the very small gaps and the leads and supports which must be located in regions where the electric field is negligible.

As secondary standards, fixed air condensers are constructed with several parallel plates and, although bulky, have very small losses.

Continuously variable condensers are usually of the parallel-plate type with a movable set of plates overlapping a fixed set by an amount determined by the rotation of a calibrated head; or of the cylindrical type, with a central conductor whose length of overlap with the outer is controlled by a micrometer screw head. The variation of capacitance in the first type can be made to obey various laws by shaping the plates, while the cylindrical type is almost linear over a wide range.

Solid- and Plastic-dielectric Condensers. Typical capacitances for air condensers lie between a few pF and about 0·01 μF, but because the plates can be much closer together solid dielectrics mean a much larger capacitance-to-volume ratio and the extremely high ε_r of some materials enhances this even further.

For relatively stable secondary standards, sheets of tinfoil and mica only a few thousandths of an inch thick are interleaved, rigidly clamped, heated to drive off moisture and then encased in wax. Mica sprayed with silver is also increasingly used in small condensers. For less exacting requirements the number of dielectric materials with special characteristics multiplies year by year. Paper, impregnated with paraffin wax, interleaved with aluminium or tin foil and rolled into a small volume, has long been the main dielectric for general use, but plastics are replacing paper and metallized films are replacing the foils. Hollow tubes of ceramic with very high ε_r can be coated with silver on the two surfaces to provide very compact condensers for miniature circuits.

Electrolytic Condensers. Electrolysis of certain salts between aluminium foils causes an insulating film of thickness about 10^{-5} cm to be formed on one foil. Used as a dielectric, this film enables capacitances of up to several thousand μF to be obtained in a reasonable volume although with a relatively low working voltage. Since quite a small potential difference applied in the reverse direction would destroy the film, the main uses for these condensers are those in which uni-directional voltages are encountered. Tantalum

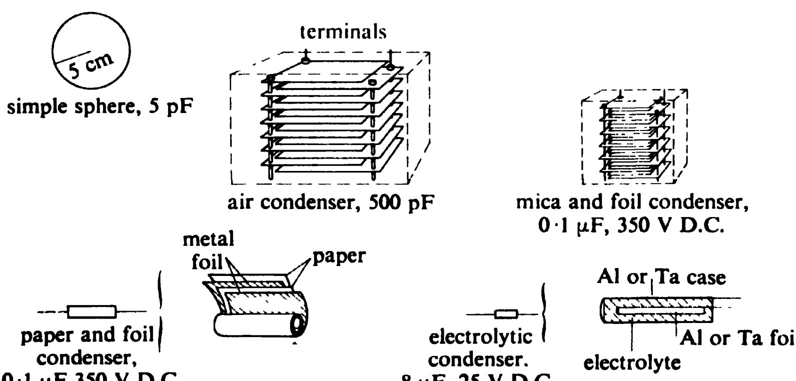


Fig. 5.14. Construction and comparative sizes of typical condensers. On this scale, some ceramic and electrolytic condensers of a few pF *would appear as small dots.*

foil is now extensively used where extreme miniaturization is required.

Figure 5.14 illustrates the construction and relative sizes of the main types of condenser.

Stray Capacitance and Screening. If the plates of a standard condenser are such that one encloses the other, the inner is completely screened from outside effects and if the outer is earthed (connected to the surroundings at constant potential) as in Fig. 5.15a, the capacitance is just that between the inner and outer. If, however, the inner plate is earthed (Fig. 5.15b), there is an additional

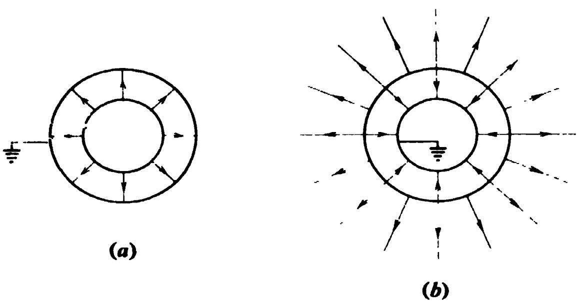


Fig. 5.15. Condenser with (a) outer conductor earthed, (b) inner conductor earthed.

capacitance between the outer and the surroundings in parallel (common V) with the original. Problem 5.8 illustrates this and shows that where the condenser takes this form, the outer should usually be earthed.

A condenser often does not have one plate screening another and there may be capacitance between both plates and earth as in Fig. 5.16a. These, known as *stray capacitances*, may be appreciable fractions of C particularly in air condensers, and they make the

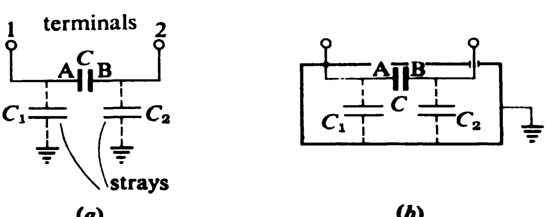


Fig. 5.16. Stray capacitances and screening.

actual capacitance between A and B indeterminate. A complicating factor in variable condensers is the change in the capacitance between the movable plate and earth as the hand approaches the knob to turn it and as the knob is turned. To avoid this, a metal case is provided to which plate A is connected as in Fig. 5.16b. C_1 is now eliminated and C_2 is quite definite so that the actual capacitance is now $C+C_2$ and this is the value quoted by the makers. If C is variable, the movable plate is connected to A.

5.5 Electric Energy 1: In Terms of Q and V

Work Done in Charging a Condenser and a Conductor. Whenever an agency such as a voltaic cell charges a condenser it does work in transferring the charge. Figure 5.17 represents the charging of a condenser of capacitance C to a final potential difference V

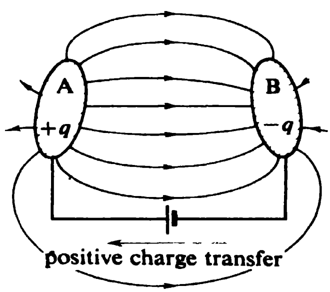


Fig. 5.17. A condenser AB *charged by a cell has charges conveyed from one plate to the other against the electrostatic field.*

produced by a total charge $\pm Q$. Suppose that at some stage in the charging process the potential difference has reached v and the charge already transferred is q. An additional charge $\mathrm{d}q$ increases v to $v+\mathrm{d}v$ and the work done by the cell lies between $v\,\mathrm{d}q$ and $(v+\mathrm{d}v)\,\mathrm{d}q$ (from the definition of V in equation (3.8): work done $=$ QV). In the limit we may neglect the second order term and write the work as $v\,\mathrm{d}q$. The total work done is thus $\int_0^Q v\,\mathrm{d}q$ and since $q=vC$ at all stages this is $\frac{1}{2}CV^2$. Hence

$$W = \tfrac{1}{2}QV = \tfrac{1}{2}CV^2 = \tfrac{1}{2}Q^2/C \tag{5.15}$$

gives the work done in charging a condenser.

We can regard a single conductor remote from its surroundings as a special case and use (5.15) with the understanding that V is the potential (zero at infinity) of the conductor.

Electric Energy of Charged Conductors. Because work is done in moving charges to their final positions in a charged condenser, there has been an increase in their potential energy of amount $\frac{1}{2}QV$ or $\frac{1}{2}Q(V_A - V_B)$ where V_A and V_B are the potentials of the plates. From this last expression we see that if the charges are assembled not by transference from one plate to another but separately brought from infinity in equal amounts the work done is the same because the energy of A is $\frac{1}{2}QV_A$ and that of B is $-\frac{1}{2}QV_B$.

These expressions can be obtained in a more general way. Equation (2.12) gave an expression for the potential energy of three charges assembled from infinity:

$$U = Q_1Q_2/4\pi\varepsilon_0 r_{12} + Q_2Q_3/4\pi\varepsilon_0 r_{23} + Q_3Q_1/4\pi\varepsilon_0 r_{31} \qquad (5.16)$$

and this could be written as

$$U = \tfrac{1}{2}Q_1\left(\frac{Q_2}{4\pi\varepsilon_0 r_{12}} + \frac{Q_3}{4\pi\varepsilon_0 r_{31}}\right) + \tfrac{1}{2}Q_2\left(\frac{Q_3}{4\pi\varepsilon_0 r_{23}} + \frac{Q_1}{4\pi\varepsilon_0 r_{12}}\right) + \tfrac{1}{2}Q_3\left(\frac{Q_1}{4\pi\varepsilon_0 r_{31}} + \frac{Q_2}{4\pi\varepsilon_0 r_{23}}\right) \qquad (5.17)$$

in which each of the terms of (5.16) has been split into two halves. Equation (5.17) can be written as

$$U = \tfrac{1}{2}Q_1V_1 + \tfrac{1}{2}Q_2V_2 + \tfrac{1}{2}Q_3V_3 \qquad (5.18)$$

in which the V's are the potentials at the point occupied by the corresponding Q due to all the *other* charges. It is clear that for any number of charges this can be generalized to

$$U = \sum_i \tfrac{1}{2}Q_iV_i \qquad (5.19)$$

Suppose now that the charges are distributed over the surfaces of N conductors A, B, C, etc., so that all those charges on conductor A occupy points at which the potential is V_A and can thus be taken out of the sum in (5.19) to form a term $\frac{1}{2}Q_AV_A$ where Q_A is the total charge on A. Repeated for all the conductors, (5.19) becomes

$$U = \sum_{A=1}^{N} \tfrac{1}{2}Q_AV_A \qquad (5.20)$$

This gives $\frac{1}{2}Q(V_A - V_B)$ for an ideal condenser.

In all cases the zero of U is taken when all the charges are remote from each other at infinity or when all conductors are uncharged.

We shall refer to U in the above expressions as *electric energy* and denote it by U_E: it is potential energy stored in the system because of the positions of the charges and the electric forces between them, and is energy *internal* to the system. (As we have seen, the potential energy of Q in an *external* field is QV not $\frac{1}{2}QV$.) When the forces are allowed to move the charges, for instance through conducting wires, the electric energy decreases at the expense of heat energy and becomes zero when the conductors equalize their potentials. Alternatively, the forces may move the conductors as a whole when the electric energy decreases due to the external work done (section 5.7).

5.6 Electric Energy 2: In Terms of E

We now transfer our attention to the electric field rather than the charges and consider a single charged conductor at a potential V (zero at infinity) carrying a charge Q so that its electric energy is $\frac{1}{2}QV$. A small part of the surface carries a charge dQ and has energy $dU_E = \frac{1}{2}V\, dQ$ associated with it. The tube of lines of force generated by dQ is shown in Fig. 5.18 intersecting two typical

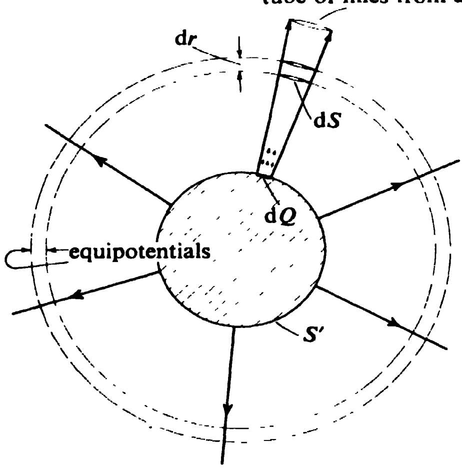


Fig. 5.18. Calculation of the energy of a charged conductor in terms of the field strength.

equipotential surfaces differing in potential by $\mathrm{d}V$ and carving out a small volume of area $\mathrm{d}S$ and thickness $\mathrm{d}r$. If the electric field at the volume element is denoted by E then Gauss's theorem, applied to the volume bounded by $\mathrm{d}Q$, $\mathrm{d}S$ and the sides of the tube, gives $E\,\mathrm{d}S=\mathrm{d}Q/\varepsilon_0$ so that

$$\mathrm{d}U_{\mathrm{E}} = \tfrac{1}{2}\varepsilon_0 E\,\mathrm{d}S\,V$$

But $V=\int_{\infty}^{S'} -E\,\mathrm{d}r=\int_{S'}^{\infty} E\,\mathrm{d}r$ and is the sum of elementary potential differences stretching from S' to infinity and, since $\mathrm{d}Q=\varepsilon_0 E\,\mathrm{d}S$ applies to all the equipotentials cut by the tube,

$$\mathrm{d}U_{\mathrm{E}} = \tfrac{1}{2}\varepsilon_0 E\,\mathrm{d}S\int_{S'}^{\infty} E\,\mathrm{d}r = \int_{S'}^{\infty} \tfrac{1}{2}\varepsilon_0 E^2\,\mathrm{d}S\,\mathrm{d}r$$

The same expression applies to every tube starting from a $\mathrm{d}Q$ and the total energy is obtained by integrating over the other two coordinates represented by $\mathrm{d}S$. Thus if the volume element $\mathrm{d}S\,\mathrm{d}r$ is denoted by $\mathrm{d}\tau$,

$$U_{\mathrm{E}} = \iiint \tfrac{1}{2}\varepsilon_0 E^2\,\mathrm{d}\tau \tag{5.21}$$

where the volume integration extends over the whole of space, including now the inside of the conductor where $E=0$ and no contribution to U_{E} is made. We have, in short, assumed that the energy $\frac{1}{2}QV$ can be divided up so that $\frac{1}{2}\,\mathrm{d}Q\,\mathrm{d}V$ is associated with the volume element $\mathrm{d}\tau$; shown that $\frac{1}{2}\,\mathrm{d}Q\,\mathrm{d}V$ can be written as $\frac{1}{2}\varepsilon_0 E^2$; and integrated over space to obtain (5.21).

This derivation can be extended to an ideal condenser in which all the tubes of lines end on another conductor instead of going to infinity, and thence to sets of charged conductors by considering all tubes starting from $\mathrm{d}Q$ on one conductor and ending at $-\mathrm{d}Q$ on another.

That the two expressions (5.20) and (5.21) give the same result for the total energy can be exemplified by a parallel-plate condenser with a charge Q and capacitance $C=\varepsilon_0 A/x$. Equation (5.21) gives U_{E} as $\frac{1}{2}Q^2x/\varepsilon_0 A$ since $E=Q/\varepsilon_0 A$ everywhere between the plates and the volume is xA. This is the same as $\frac{1}{2}Q^2/C$ or $\frac{1}{2}QV$.

Equation (5.21) can be interpreted by saying that the energy per unit volume, or the energy density, is $\frac{1}{2}\varepsilon_0 E^2$ but we shall not take the view that this implies any storage of energy in space since we must always use the integrated expression giving the total energy of

the system *as a whole*: there is little point in arguing about the location of the energy.

5.7 Forces, Couples and Changes in Energy

The expressions (5.20) and (5.21) for the internal potential energy of charged conductors and condensers can be used to obtain the forces and couples acting on any particular conductor. The method (of allowing the conductor to move under the force) is similar to that used in sections 3.5 and 3.6 to obtain **E** from V and leads to a similar result. Here, however, the movement may take place with the conductors isolated (Q constant) or connected to batteries (V constant) and both cases must be considered.

Changes at Constant Q. Consider the two conductors of a condenser as in Fig. 5.19 where the internal forces of attraction are shown dotted and the external forces keeping the plates apart and in

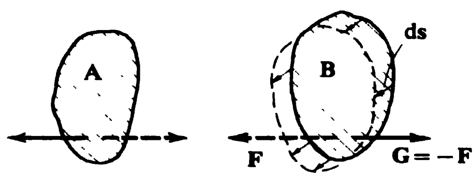


Fig. 5.19. Forces between charged conductors are dotted. The external forces maintaining equilibrium are full.

equilibrium are shown full. Conductor B has an internal force **F** and an external force $\mathbf{G} = -\mathbf{F}$ acting on it. For any displacement $\mathbf{ds}$ of B the increment of internal energy is dU_E and the work done *on* the condenser *by* **G** is $\mathbf{G}\cdot\mathbf{ds}$. Hence

$$dU_E = \mathbf{G}\cdot\mathbf{ds} = -\mathbf{F}\cdot\mathbf{ds} = -F_s\,ds$$

or

$$F_s = -\left(\frac{\partial U_E}{\partial s}\right)_Q \tag{5.22}$$

(cf. $E_s = -\partial V/\partial s$) giving the component of the force on B in any direction **s**. Because the charges are constant

$$dU_E = \tfrac{1}{2}Q\,dV = \tfrac{1}{2}Q^2\,d(1/C) = -\tfrac{1}{2}V^2\,dC$$

so that from (5.22)

$$F_s = \tfrac{1}{2}V^2\frac{\partial C}{\partial s} \tag{5.23}$$

Similarly, the couple about any axis is given by

$$T_\theta = -\left(\frac{\partial U_E}{\partial \theta}\right)_Q \tag{5.24}$$

which gives, for one conductor of a condenser

$$T_\theta = \tfrac{1}{2}V^2 \frac{\partial C}{\partial \theta} \tag{5.25}$$

As a simple example, the force between the plates of a parallel-plate condenser of $C = \varepsilon_0 A/x$ and carrying a charge density σ would be

$$\begin{aligned} F_x = \tfrac{1}{2}V^2 \frac{dC}{dx} &= -\tfrac{1}{2}\varepsilon_0 A \frac{V^2}{x^2} \\ &= -\tfrac{1}{2}\varepsilon_0 E^2 A \quad \text{or} \quad -\sigma^2 A/2\varepsilon_0 \end{aligned} \tag{5.26}$$

The sign indicates that F_x acts to decrease x: it is an attraction. The reader should confirm that (5.26) can also be obtained from (2.7)—a longer method—or from (5.22), a shorter one.

Changes at Constant V. Suppose now that the same condenser is maintained by a battery at a constant potential difference V as in Fig. 5.20, and that conductor B again undergoes a displacement ds.

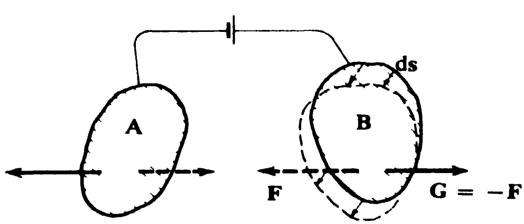


***Fig. 5.20.** Displacement of the conductor in Fig. 5.19 at constant potential.*

Let the increment of potential energy again be dU_E, this time equal to $\tfrac{1}{2}V\,dQ$ and not to $\tfrac{1}{2}Q\,dV$. The battery has to transfer a charge dQ to keep V constant and in doing so does work $V\,dQ$ and thus loses this amount of energy. The increment in the energy of the system (condenser + battery) is thus $-V\,dQ + \tfrac{1}{2}V\,dQ$ or $-\tfrac{1}{2}V\,dQ$ which equals $-dU_E$. As before, the work done *on* the condenser is

G·d**s** or −**F**·d**s** and so

$$-\mathrm{d}U_{\mathrm{E}} = \mathbf{G}\cdot\mathrm{d}\mathbf{s} = -\mathbf{F}\cdot\mathrm{d}\mathbf{s} = -F_s\,\mathrm{d}s$$

or

$$F_s = +\left(\frac{\partial U_{\mathrm{E}}}{\partial s}\right)_V \tag{5.27}$$

Because V is constant, this gives $F_s = \frac{1}{2}V^2(\partial C/\partial s)$ as in (5.23): naturally, however the *changes* are made, the forces in a static situation are the same.

Notice that changes at constant V mean that of the energy supplied by the batteries, half goes to increasing the electric energy U_{E} and half goes in doing external work.

Force on the Surface of a Charged Conductor. Of the several methods of calculating this force we choose two: one to show some detail of how the force arises and the other to illustrate the power of the energy method.

Figure 5.21 represents a part of the surface of a charged conductor where the surface density of charge is σ. This charge resides in a

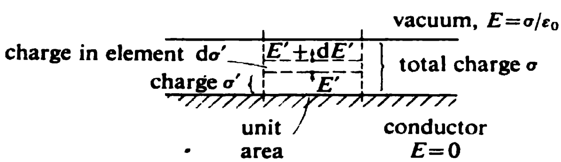


Fig. 5.21. ***Detail of the surface of a charged conductor.***

thin surface layer of unspecified thickness. Consider unit area of the layer containing a total charge σ, spread in fact throughout a cylinder, at the bottom of which $E=0$ and at the top of which $E=\sigma/\varepsilon_0$. Since E must increase through the layer let its magnitude at some level be E' and the charge in the cylinder up to this level be σ'. A small element of the cylinder containing a charge $\mathrm{d}\sigma'$ thus experiences an outward force $E'\,\mathrm{d}\sigma'$. If, however, the electric field increases across the element to $E'+\mathrm{d}E'$ the application of Gauss's theorem to the same element gives $\mathrm{d}E'=\mathrm{d}\sigma'/\varepsilon_0$ and the force on the element is this $\varepsilon_0 E'\,\mathrm{d}E'$. On the whole cylinder the force is

$$\int_0^E \varepsilon_0 E'\,\mathrm{d}E' \text{ which is } \tfrac{1}{2}\varepsilon_0 E^2 \text{ or } \sigma^2/2\varepsilon_0.$$

This force is strictly exerted on the charges in the surface, but because emission of charge does not normally occur, there must be a

balancing non-electrostatic force exerted by the conductor on the charges. This causes the force to be communicated to the conductor itself by Newton's third law so that

Outward pressure on the surface of a charged conductor

$$= \frac{\varepsilon_0 E^2}{2} = \frac{\sigma^2}{2\varepsilon_0} \qquad (5.28)$$

At sharp points, where σ is very large, the force *can* be great enough to cause charges to leave the surface and *field emission* is then said to occur.

Using (5.21) we can also find the force on the surface by allowing a virtual displacement dx of area A to occur under the action of the outward force F_x. By equating the work done by the internal force, F_x dx, to the decrease of energy $U_E A$ dx because of the disappearance of a volume A dx in which E existed, (5.28) is again obtained.

5.8 Electrostatic Instruments and Measurements

Electrostatic instruments contain two or three conductors, one of which is movable under the action of forces or couples given by (5.23) or (5.25): they thus possess an inherent capacitance and are essentially measurers of potential difference. They are in general known as *electrometers* although when only detection of charge or p.d. is required the term *electroscope* is more appropriate (section 1.1). *Electrostatic voltmeter*, a term sometimes used for electrometers measuring upwards of about a volt, will here denote only the particular instrument described as such below.

Electrometers can be used to measure small currents in one of two ways:

(*a*) the current is passed through a high resistance and the potential difference developed across this is measured;

(*b*) the current is passed straight to the electrometer, which charges, or it is allowed to discharge a charged instrument: the current is related to the rate of change of deflection, dθ/dt.

Electromagnetic instruments, by contrast, are essentially current measurers but can be adapted to measure potential difference (section 16.5).

If the voltage sensitivity (deflection per unit potential difference) is S_V then

$$S_V = \frac{d\theta}{dV} \qquad (5.29)$$

If method (*a*) above is used for current measurement, the current sensitivity S_I or $\mathrm{d}\theta/\mathrm{d}I$ is $\mathrm{d}\theta/\mathrm{d}(V/R)$ by (6.8) so that

$$S_I = RS_V$$

For method (*b*), the current $I = \mathrm{d}Q/\mathrm{d}t = C\,\mathrm{d}V/\mathrm{d}t$ if C is constant. Thus

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\mathrm{d}\theta}{\mathrm{d}V} \times \frac{\mathrm{d}V}{\mathrm{d}t} = \frac{S_V}{C} I$$

For charge measurement, the quantity or charge sensitivity S_Q or $\mathrm{d}\theta/\mathrm{d}Q$ is

$$S_Q = S_V/C$$

It follows that measuring current by method (*a*) requires a high resistance while measuring current by method (*b*) or measuring a charge requires a low capacitance instrument.

Two-electrode Instruments. In these, the force or couple given by (5.23) or (5.25) is very small unless large potential differences are used, but the dependence on V^2 means that the polarity of the potential difference is immaterial so that alternating voltages may be measured.

One class of two-electrode instrument consists of various modifications of electroscopes whose capacitances are so low (a few pF) that they have high S_I and S_Q. They are chiefly used to measure ionization currents although they will operate as crude voltmeters. The instrument of section 1.1 is, we now see, a measurer of potential difference between leaf and case, but it can be used as a charge measurer since $Q \propto V$ for a constant C, and the variation of C as the leaf or needle deflects is negligible compared with the magnitude of C itself. Examples of these instruments are the *Braun electrometer*, an electroscope like that of Fig. 1.1 but with a light aluminium needle and a scale which can be used for rough measurements from 500 V to 100 kV; and the *Lauritsen electrometer* shown in Fig. 5.22

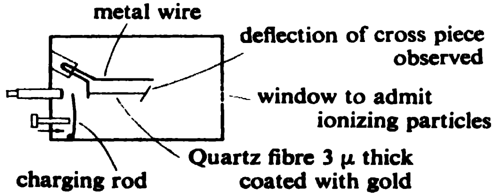


Fig. 5.22. The Lauritsen electrometer.

which has a capacitance of only about 0·2 pF. A compact form of the Lauritsen electrometer is now used as a pocket dosemeter for radiation: a standard deflection is given to the instrument by charging it and the decay over a given time indicates the total ionization charge received. Ionization currents are now measured principally by either of methods (*a*) or (*b*) above but using the grid-cathode capacitance of a specially-designed *electrometer valve* as the element across which the potential difference is developed: these do not operate as a result of forces and should properly be classed as electronic instruments.

The second class of two-electrode instruments consists of those used for direct measurement of potential difference. The *attracted disc electrometer* (Fig. 5.23) does not need calibration if ε_0 is known

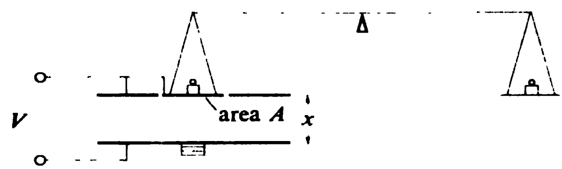


Fig. 5.23. The attracted disc electrometer.

because the force is accurately $\varepsilon_0 A V^2/2x^2$ by (5.26) owing to the presence of guard-rings. The guarded ring is balanced with $V=0$ and a weight mg on it. This weight is then removed and the ring rebalanced by applying V and adjusting either it or the distance x. The method is tedious and not suited to routine measurement: it has been used recently at the U.S. National Bureau of Standards for measuring alternating potential differences of up to 250 kV in order to check the accuracy of routine methods normally used.

In the *electrostatic voltmeter* (Fig. 5.24), the couple on the vane,

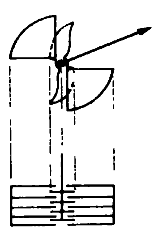


Fig. 5.24. An electrostatic voltmeter.

$\frac{1}{2}V^2\,\mathrm{d}C/\mathrm{d}\theta$, is opposed by that due to a spring of torsional constant c so that at equilibrium

$$\theta = \tfrac{1}{2}\frac{V^2}{c}\frac{\mathrm{d}C}{\mathrm{d}\theta}$$

and the way in which θ varies with V depends on the shape of the vanes. Electrostatic voltmeters are frequently used for measuring alternating voltages and the U.K. National Physical Laboratory uses one as a transfer standard D.C. to A.C. (see section 16.4).

Three-electrode Instruments. In these, the movable conductor at potential V has a capacitance C_1 with one stationary conductor at a potential V_1, and C_2 with another at V_2. Thus the total torque is

$$T_\theta = \tfrac{1}{2}\frac{\mathrm{d}C_1}{\mathrm{d}\theta}(V-V_1)^2+\tfrac{1}{2}\frac{\mathrm{d}C_2}{\mathrm{d}\theta}(V-V_2)^2$$

assuming that the capacitance between the fixed conductors does not change. T_θ is balanced by the torsional couple $c\theta$ of a fibre in the *quadrant electrometer* and the *Lindemann electrometer* (Figs. 5.25 and 5.26). Since in these the vane or needle is symmetrical

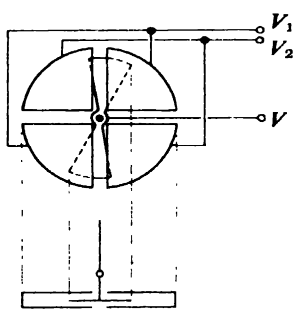


Fig. 5.25. The quadrant electrometer.

$\mathrm{d}C_1/\mathrm{d}\theta = -\mathrm{d}C_2/\mathrm{d}\theta$, and so

$$T_\theta = \tfrac{1}{2}\frac{\mathrm{d}C}{\mathrm{d}\theta}(2V-V_1-V_2)(V_2-V_1)$$

$$\doteqdot V\frac{\mathrm{d}C}{\mathrm{d}\theta}(V_2-V_1)$$

if V is made much greater than V_1 or V_2. The potential difference

to be measured is $(V_2 - V_1)$ and if, as in the quadrant electrometer, C is proportional to θ, $dC/d\theta$ is a constant K and $c\theta = KV(V_2 - V_1)$.

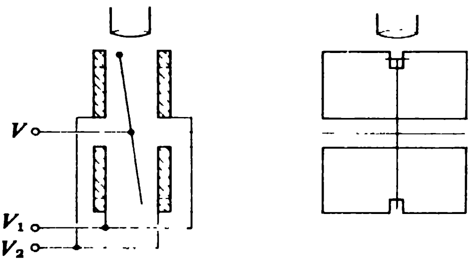


Fig. 5.26. The Lindemann electrometer.

The voltage sensitivity S_V is thus KV/c which can be increased by using higher vane or needle potentials. A similar analysis applies to the *quartz fibre electrometer* shown in Fig. 5.27.

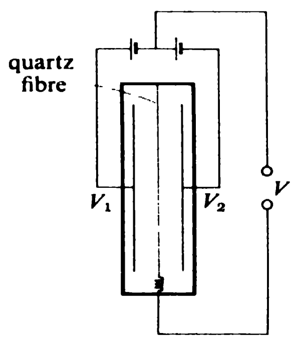


Fig. 5.27. The quartz fibre electrometer.

Electrostatic Measurements. Few measurements are now carried out using electrostatic methods. Potential difference is normally measured by methods described in chapter 6, amongst which is the electrometer method: electric field strength is obtained from potential gradient; capacitance is usually obtained from an A.C. bridge method (section 10.12); and charge by ballistic galvanometer (section 16.6). Nevertheless an electrostatic instrument should always be considered when measurements are to be made involving very high potentials or negligible current consumption or

which must be independent of whether the potential is alternating or steady.

The Value of ε_0. While we shall be dealing with the accurate measurement of ε_0 in section 16.7 we need at this stage to have a value for it which we can use. Eichenwald's experiment in section 1.3 in fact yielded a value (although he would have regarded it as a value for the ratio of the e.m.u. of charge to the e.s.u.—see section 16.7). In table 1.1 the column headed Q was in reality the potential in volts although, because of the constancy of the capacitance, this provided an arbitrary measure of charge which was all we needed at that stage. The column headed I was the current in μA. If the capacitance of the rotating disc, etc. was C, then in the experiment $nVC = I$ or $C = I/nV$. From the last column the mean value of nV/I was $2{\cdot}18 \times 10^{10}$ so that C was 45·9 pF. Using the dimensions and a value for $4\pi\varepsilon_0$ of $1/(9 \times 10^9)$, Eichenwald found $C = 45{\cdot}6$ pF. This justifies with sufficient accuracy for now our value in equation (2.6).

5.9 Electrostatic Generators, Hazards, etc.

The Van de Graaff generator illustrated in Fig. 5.28 is used to produce beams of ions with a very small spread of energy. If positive ions are to be accelerated from a source S, then G maintains a comb of points A at a positive potential of between 10 and 50 kV in the larger machines, and charge is sprayed on to the moving belt of insulating material by corona discharge. The belt carries the charge towards the upper pulley inside the dome and here the

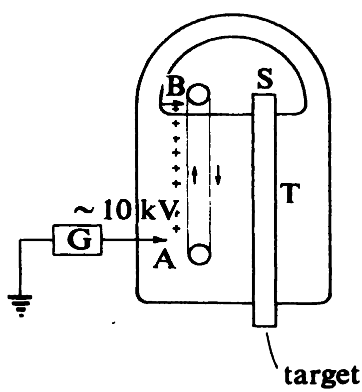


Fig. 5.28. The Van de Graaff generator.

points of comb B become negatively charged by induction. Corona again occurs and positive charge is transferred to the outside of the dome, the belt being discharged. The potential of the dome and its interior thus rises until the supply of charge from the belt is balanced by the currents due to the accelerated ions and the leakage through the gas surrounding the dome and the tube T. Machines producing ions with energies above 5 MeV are in use, and the energy can be at least doubled by using a *tandem generator* in which negative ions are accelerated towards the dome, stripped of some electrons within it so that they become positive ions, and then again accelerated but this time away from the dome (Allen, 1920).

What is, at first sight, another type of generator altogether is based on the principle of the condensing electroscope. In Fig. 5.29, two conductors A and B are very close and have a capacitance C_i between them. A is earthed and B is raised to a small potential

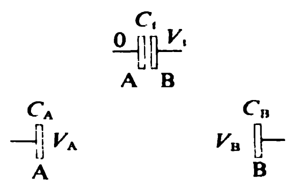


Fig. 5.29. The principle of the condensing electroscope.

V_i so that the magnitude of the charge on both is given by $Q = V_iC_i$. They are then isolated and separated by such a large distance that they both function almost as isolated conductors with capacitances C_A and C_B with respect to the surroundings. Because they retain the same charge, the final potential of B is V_iC_i/C_B and of A is V_iC_i/C_A. Since C_i is much greater than either C_A or C_B, the final potential of both conductors is much higher than the original, one being negative and the other positive. The condensing electroscope itself, the electrophorus and charging by induction are all examples of the process described in elementary texts, although the importance of the increased potential is often not pointed out.

Continuous generators using the principle have recently been developed at Grenoble (Bright, 1923). A rotor has vanes which pass between a pair of plates at a slightly higher potential and each vane is earthed as it passes, the connection being broken almost at once. When the vane has moved so far that the capacitance is a minimum,

and the potential therefore a maximum, the charge is collected. Although the Van de Graaff generator is generally regarded as operating on a different principle, there is evidence that it too works by the condensing electroscope mechanism (Simon, 1922; Subudhi and Tiwari, 1922). Experiments show that the potential of the dome rises exponentially instead of linearly to a limit which depends on the spraying voltage as well as on the insulation and these results support the idea that the high potential developed is due not to the accumulation of charge in a Faraday ice-pail but to the work done against electric fields.

Generation of electric charge is often a hazard or at best a nuisance in industrial processes, although it can be put to use in the electrostatic precipitation of dusts. References to these and other aspects of electrostatics are given at the end of the chapter.

5.10 Coefficients of Potential, Capacitance and Induction

Most of the practical situations met in this chapter have concerned not more than two conductors and the surroundings. A method for dealing with any finite system of conductors can be obtained by a generalization of the concepts introduced so far. To avoid getting lost in a turmoil of subscripts, we shall use 3 conductors only and make the obvious extensions.

Suppose the conductors in Fig. 5.30 possess charges Q_1, Q_2 and Q_3 and are thereby raised to potentials V_1, V_2 and V_3 with respect to the surroundings. Let

p_{11} = potential of 1 due to unit charge on 1, zero on 2 and 3 (5.30)

p_{12} = potential of 1 due to unit charge on 2, zero on 3 and 1 (5.31)

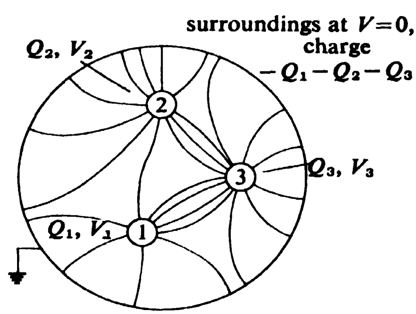


Fig. 5.30. Lines of force of three charged conductors in earthed surroundings.

and similarly for p_{13}, p_{22}, etc., noting that p_{11} is produced by induced charges on 2 and 3 as well as by the placed charge on 1. The argument of section 5.1 applied to all surfaces shows that Q and V are proportional and hence that

$$\left.\begin{aligned} p_{11}Q_1 &= \text{potential of 1 due to } Q_1 \text{ on 1, zero on 2 and 3} \\ p_{12}Q_2 &= \text{potential of 1 due to } Q_2 \text{ on 2, zero on 3 and 1} \\ p_{13}Q_3 &= \text{potential of 1 due to } Q_3 \text{ on 3, zero on 1 and 2} \end{aligned}\right\} \tag{5.32}$$

By superposition, when Q_1 is placed on 1, Q_2 on 2 and Q_3 on 3

$$V_1 = p_{11}Q_1 + p_{12}Q_2 + p_{13}Q_3$$

and similarly

$$\left.\begin{aligned} V_2 &= p_{21}Q_1 + p_{22}Q_2 + p_{23}Q_3 \\ V_3 &= p_{31}Q_1 + p_{32}Q_2 + p_{33}Q_3 \end{aligned}\right\} \tag{5.33}$$

The p's are known as coefficients of potential and, generalizing to N conductors, we can write

$$V_i = \sum_{j=1}^{N} p_{ij}Q_j \qquad \text{for } i = 1 \text{ to } i = N \tag{5.34}$$

Equations (5.33) can be solved for the Q's and written in the form

$$Q_1 = c_{11}V_1 + c_{12}V_2 + c_{13}V_3, \text{ etc.} \tag{5.35}$$

or in general

$$Q_i = \sum_{j=1}^{N} c_{ij}V_j \qquad \text{for } i = 1 \text{ to } i = N \tag{5.36}$$

in which the c_{ii} are known as coefficients of capacitance and the c_{ij} as coefficients of induction. If Δ is the determinant of the p_{ij}'s

$$c_{ij} = (\text{cofactor of } p_{ij} \text{ in } \Delta)/\Delta \tag{5.37}$$

from (5.34) and (5.36).

Energy and Forces. From (5.20), the internal electric energy of a system of charged conductors is

$$U_E^{\bullet} = \tfrac{1}{2}\sum_i \sum_j p_{ij}Q_iQ_j = \tfrac{1}{2}\sum_i \sum_j c_{ij}V_iV_j \tag{5.38}$$

while the force on the ith conductor by methods similar to those already used is

$$F_{s_i} = -\left(\frac{\partial U_E}{\partial s_i}\right)_Q = -\tfrac{1}{2}\sum_i\sum_j Q_iQ_j\frac{\partial p_{ij}}{\partial s_i}$$

or

$$F_{s_i} = +\left(\frac{\partial U_E}{\partial s_i}\right)_V = +\tfrac{1}{2}\sum_i\sum_j V_iV_j\frac{\partial c_{ij}}{\partial s_i}$$

Link with Previous Work. Reciprocity. The system of 3 conductors as in Fig. 5.30 have capacitances with each other and with the surroundings as shown in Fig. 5.31. Using the definitions of

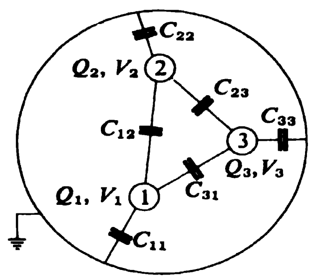


Fig. 5.31. Capacitances between conductors and surroundings in Fig. 5.30.

capacitance, C_{11} means the ratio of Q_1 to V_1 when no influences from 2 and 3 affect 1, i.e. 1, 2 and 3 are all connected together. C_{12} means the ratio of Q_1 to $(V_1 - V_2)$ when no influences other than 2 affect 1, i.e. 1 and 3 are earthed, and hence $V_1 = V_3 = 0$. Using (5.35),

$$C_{11} = c_{11}+c_{12}+c_{13}$$
$$C_{12} = -c_{12}$$

with similar relations between the other C's and c's. Since the C's are essentially positive, all c_{ij} must be negative and c_{ii} positive. Moreover, since both c_{12} and c_{21} are equal to $-C_{12}$ we shall have in general that $c_{ij}=c_{ji}$ and hence by (5.37) that $p_{ij}=p_{ji}$.

If values of the p's or c's are required in any particular example, they can be obtained by choosing specific values for the Q's say and finding the resultant V's, since all the coefficients are independent of Q and V. (See problems 5.22 and 5.23.)

5.11 Summary of Chapter 5

In progressing from one charged conductor to a system of conductors we have introduced and generalized the concept of capacitance in sections 5.1, 5.2, 5.3 and 5.10. In the absence of dielectrics we have shown that capacitances are constant for a given geometry if the electrostatic laws apply. We have also derived some general formulae for the internal electric energy of a system of charges or conductors

$$U_E = \tfrac{1}{2}\sum_i Q_i V_i = \iiint \tfrac{1}{2}\varepsilon_0 E^2\, d\tau$$

and have used these to calculate forces and couples and to examine the principles of electrostatic instruments.

Once again, most of the results in this chapter are consequences of Coulomb's law: the only exception to this lies in the introduction of dielectric materials, whose properties can only be discovered by further experiments. Readers who wish to carry straight on to a study of these may proceed directly to chapter 13, the initial parts of which can be assimilated without trouble at this stage.

References

Types and construction of condensers: Dummer (1959). Electrostatic instruments: Harris (1952), Estermann (1959). Electrostatic generators, hazards, etc.: *British Journal of Applied Physics* (1953), Harper (1961).

PROBLEMS

SECTION 5.1

5.1 Use the formula for the capacitance of a sphere to find the relation between 1 e.s.u. of capacitance and 1 pF.

SECTION 5.2

*5.2 Using the method at the end of section 5.2, find the limits between which the capacitance of a conducting cylinder of length $2l$ and radius a must lie. As the length is increased relative to the diameter, to what value will the capacitance per unit length at the centre tend?

5.3 Estimate the capacitance of (*a*) a halfpenny, (*b*) a man, considered as an isolated conducting cylinder of length 2 m and diameter 30 cm, (*c*) a straight copper wire of length 10 cm, diameter 1 mm and situated 1 cm away from and parallel to a plane earthed chassis, (*d*) the earth.

*5.4 A twin transmission line consists of two thin conducting strips of width $2a$ with their flat surfaces facing each other and separated by a distance x. Find the capacitances per unit length on the extreme assumptions of uniform distribution of charge and distribution only along the edges.

*5.5 Find, using the solution of problem 5.4, whether the edge effects increase or decrease the capacitance of a parallel-plate condenser compared with its ideal value.

SECTION 5.3

5.6 Condensers A and B of 0·5 μF and 0·2 μF respectively are connected in series across a 140 V D.C. supply. What are the charges and potential differences for each condenser? Repeat with A and B in parallel.

5.7 Show that, if the arrangement of Fig. 5.32a is to be equivalent to that of Fig. 5.32b when the points A, B and C are connected into any network,

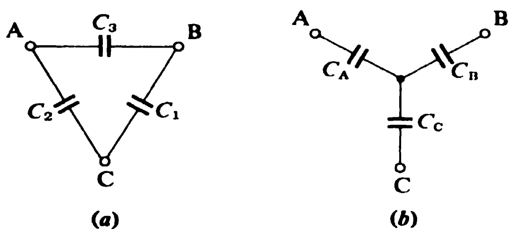


Fig. 5.32. The delta-Y transformation.

$C_A = (C_1C_2 + C_2C_3 + C_3C_1)/C_1$ with corresponding expressions for C_B and C_C. (Use $K_A = 1/C_A$.) Show also that $C_1 = C_BC_C/(C_A + C_B + C_C)$ etc. by using $C_1C_A = C_2C_B = C_3C_C = k$ say. Solve the network of Fig. 5.12 by reducing one of the deltas to a Y. (The equivalence of the networks of Fig. 5.32 is known as the delta-Y or π-T transformation.)

5.8 If the inner sphere of a spherical condenser is earthed instead of the outer, show that the total capacitance is $4\pi\varepsilon_0 b^2/(b-a)$ where $a<b$. If a charge Q is given to the outer sphere from surroundings at earth potential, what proportions reside on the outer and inner surfaces of the outer sphere?

SECTION 5.4

5.9 Estimate the capacitance of the following: (a) an air condenser of 16 plates each 6 cm × 4 cm separated by 2 mm, alternate plates being connected, (b) a rolled paper condenser with strips of tin-foil 3 cm × 40 cm as the plates, the paper being 0·02 mm thick and having $\varepsilon_r = 2$, (c) a ceramic condenser with plates 1 cm square and a dielectric of thickness 0·5 mm and $\varepsilon_r = 3{,}000$.

5.10 If the largest safe potential gradient which can be allowed in air is 3 kV/mm, what is the greatest capacitance which can be obtained in a volume 10 cm × 8 cm × 6 cm using air as a dielectric and having a working potential of 600 V?

SECTION 5.5

5.11 In problem 5.6, compare the energy stored in the condensers with that supplied by the source.

5.12 Eight identical spherical drops of mercury charged to 12 V above earth potential are made to coalesce into a single spherical drop. What is the new potential and how has the internal electric energy of the system changed?

5.13 A condenser of capacitance C_1 and potential difference V is connected across an uncharged condenser of capacitance C_2. Show that the new potential difference is $VC_1/(C_1+C_2)$ and that the loss in electric energy is $\frac{1}{2}C_1C_2V^2/(C_1+C_2)$.

SECTION 5.6

5.14 Show that the electric energy of a charged conducting sphere of radius a carrying a charge Q is $Q^2/8\pi\varepsilon_0 a$. Carry out the calculation by two methods, using first equation (5.20) and then equation (5.21).

5.15 Show that the electric energy of a spherical region of space of radius a filled with a charge Q spread uniformly throughout it is $3Q^2/20\pi\varepsilon_0 a$.

SECTION 5.7

5.16 Find the mechanical work needed to double the separation of the plates of a parallel-plate condenser *in vacuo* if a battery maintains them at a constant potential difference V and the area and original separation are A and x respectively.

5.17 The central conductor of a cylindrical condenser has only half its length overlapping the outer. What is the force on the inner conductor if the potential difference is V and the radii are a and b?

SECTION 5.8

5.18 The plates of a parallel-plate condenser are separated by a distance x_1. The condenser is charged and the plates isolated. If one plate is vibrated sinusoidally with an angular frequency ω and an amplitude x_0 normal to its plane, show that the potential difference is of the form $V_1+V_0 \sin \omega t$, where $V_0/V_1 = x_0/x_1$. (Vibrating reed electrometer.)

SECTION 5.9

5.19 Explain the following: (*a*) as a highly-charged body is gradually moved nearer the cap of an already-charged electroscope, the deflection first decreases to zero and then increases; (*b*) as a highly-charged object is gradually moved nearer a suspended body, the latter is at first repelled and then attracted.

5.20 In Fig. 5.33, find the output potential difference after many reversals of the ganged switches (principle of the Cockcroft-Walton generator).

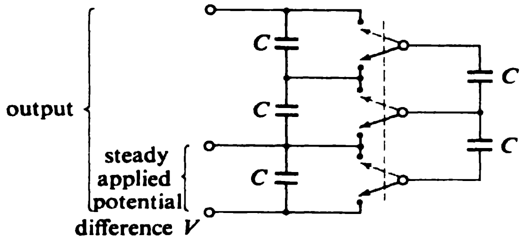


Fig. 5.33. Network for problem 5.20.

5.21 In Fig. 5.34, find the output potential difference from AB if the switches are simultaneously changed from position 1 to position 2, it being assumed that the generator of the input voltage is first isolated (principle of the cascade generator).

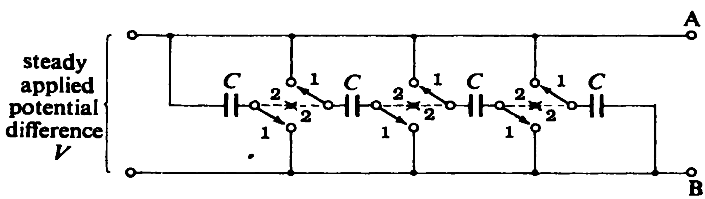


Fig. 5.34. Network for problem 5.21.

SECTION 5.10

5.22 Show that for an ideal parallel-plate condenser the coefficients of capacitance and induction are all $\pm\varepsilon_0 A/x$.

5.23 If a conductor B is connected to earth and encloses another conductor A, show, by using the coefficients of potential, that a third conductor C outside B can in no way affect the potential of A, and vice versa.

CHAPTER 6

STEADY ELECTRIC CURRENTS

In chapter 1 we saw that temporary currents could be produced in conductors by electrostatic fields but that the maintenance of a steady current required a complete conducting circuit incorporating a source of non-electrostatic field or electromotive force. We begin in section 6.1 by making more precise this concept of e.m.f., but in the rest of the chapter our interest will lie mainly in the relation between the electric field or potential difference producing a current and the current itself, a relation expressed in terms of resistance, conductance, resistivity or conductivity (sections 6.2, 6.6). Resistance, in a similar way to capacitance, is embodied in resistors which combine together and with sources of e.m.f. to give networks (sections 6.3, 6.4, 6.5) whose properties are important because so many measurements depend on them (sections 6.7, 6.8). Finally, we look briefly at evidence for the nature of the carriers of the current in conductors.

6.1 Electromotive Force

The electric fields considered in detail so far have been electrostatic. For these, distinguished where necessary by $\mathbf{E}_Q$, we have

$$\oint_C \mathbf{E}_Q \cdot d\mathbf{s} = 0 \tag{6.1}$$

for any closed path C. Charges at rest may, however, be moved by forces other than electrostatic ones as, for example, in the Van de Graaff generator. These forces constitute an electric field according to definition (3.1) and such fields will be denoted by $\mathbf{E}_M$. Their origin need not concern us at the moment but they will not in general obey (6.1), and we shall define

$$\mathscr{E} = \oint_C \mathbf{E}_M \cdot d\mathbf{s} \qquad \text{(Definition of } \mathscr{E}\text{)} \tag{6.2}$$

as the *electromotive force* or *e.m.f.* round the path C. The e.m.f. in

a closed path is therefore equal to the work done per unit positive charge in taking charge completely round the path and is measured in volts. Fields obeying (6.2) do not give rise to path-independent potential differences and we reserve the term *potential difference* specifically for electrostatic fields.

In practice it may not be possible to distinguish $\mathbf{E}_Q$ from $\mathbf{E}_M$ in a total electric field $\mathbf{E}=\mathbf{E}_M+\mathbf{E}_Q$ at any point, but if (6.1) and (6.2) are added we obtain

$$\mathscr{E} = \oint_C \mathbf{E}\cdot d\mathbf{s} \tag{6.3}$$

so that e.m.f. can be defined in terms of the total field as well.

What interests us now is the way in which e.m.f.s are accompanied by currents when the closed path C is conducting. Here, cases in which the e.m.f. is located in a part of the circuit (as with a voltaic cell) must be distinguished from those in which it extends round the whole circuit (as with some electromagnetically induced e.m.f.s—chapter 9).

Localized E.m.f.s. Figure 6.1a represents part of a circuit, the source of e.m.f., in which a field $\mathbf{E}_M$ exists. By itself, $\mathbf{E}_M$ separates

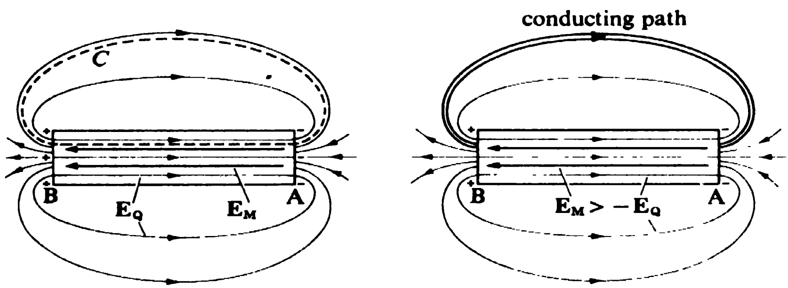


Fig. 6.1. Source of e.m.f. (a) on open circuit, (b) on closed circuit.

charges until the electrostatic field of these, $\mathbf{E}_Q$, just balances $\mathbf{E}_M$: then $\mathbf{E}_Q+\mathbf{E}_M=0$ inside the source and no further separation takes place. The source is now said to be on open circuit, its ends or terminals are charged and the e.m.f. round such a path as C is, by (6.2)

$$\mathscr{E} = \oint_C \mathbf{E}_M\cdot d\mathbf{s} = \int_A^B \mathbf{E}_M\cdot d\mathbf{s} \tag{6.4}$$

since $\mathbf{E}_M$, unlike $\mathbf{E}_Q$, does not exist outside the source. But since between A and B, $\mathbf{E}_M = -\mathbf{E}_Q$,

$$\mathscr{E} = \int_A^B -\mathbf{E}_Q \cdot d\mathbf{s} = V_B - V_A \qquad \text{(open circuit)} \qquad (6.5)$$

Thus the e.m.f. is equal to the potential difference on open circuit.

If now the ends of the source are connected by conducting wire (Fig. 6.1b) the field $\mathbf{E}_Q$ existing inside the wire causes a current to flow, the charges on the terminals being replenished by $\mathbf{E}_M$. As long as $\mathbf{E}_M$ is constant, a steady current will flow, but although this seems to imply that $\mathbf{E}_M + \mathbf{E}_Q$ would be zero in the source, $\mathbf{E}_M$ now has to work to overcome the resistance to motion of the charge so that in a closed circuit $\mathbf{E}_M > -\mathbf{E}_Q$ and hence $\mathscr{E} > V_B - V_A$ ($\mathscr{E}$ is the only e.m.f. acting).

Non-localized E.m.f.s. If the source of e.m.f. extends over the whole of a homogeneous conducting circuit then, although charge is moved by $\mathbf{E}_M$, none accumulates, no $\mathbf{E}_Q$ or potential difference arises, and the whole of the work done by $\mathbf{E}_M$ is against the resistance of the conductor to motion of the charge.

Work and Energy. Whether localized or not, an e.m.f. causing a charge Q to cross any cross-section of a circuit does work $Q\mathscr{E}$ (from the definition of $\mathscr{E}$) so that

$$\text{Work done by a source } \mathscr{E} \text{ supplying } Q = Q\mathscr{E} \qquad (6.6)$$

and

$$\text{Rate of working of a source } \mathscr{E} \text{ supplying a current } I = I\mathscr{E} \qquad (6.7)$$

Equation (6.6) also gives the consumption of energy from the source and (6.7) the rate of consumption of energy.

A source of e.m.f. thus involves the use of non-electric sources of energy to move charges. These sources will be surveyed in chapter 12: meanwhile in this chapter we shall assume that they are steady and localized.

6.2 Resistance and Conductance

A steady potential difference maintained across the ends of a conductor will maintain a steady current in it. The direction of motion of the charges depends on their sign but this is for the moment immaterial and we shall assume the conventional flow of

positive charge from high to low potential. The ratio of the potential difference, V, across the ends of the conductor to the current I through it is defined as the *resistance of the conductor*, R, and its reciprocal as the conductance G:

$$R = V/I \qquad \text{(Definition of } R\text{)} \quad (6.8)$$

$$G = 1/R = I/V \qquad \text{(Definition of } G\text{)} \quad (6.9)$$

the unit of resistance, the V/A, being the *ohm* (symbol Ω) and of conductance the *reciprocal ohm* or *mho*. A component embodying resistance is known as a *resistor* represented in part of a network as in Figs. 6.3 onwards and crossed by an arrow if variable.

With capacitance we were able to deduce constancy in the absence of dielectrics from Coulomb's law, but this is not possible

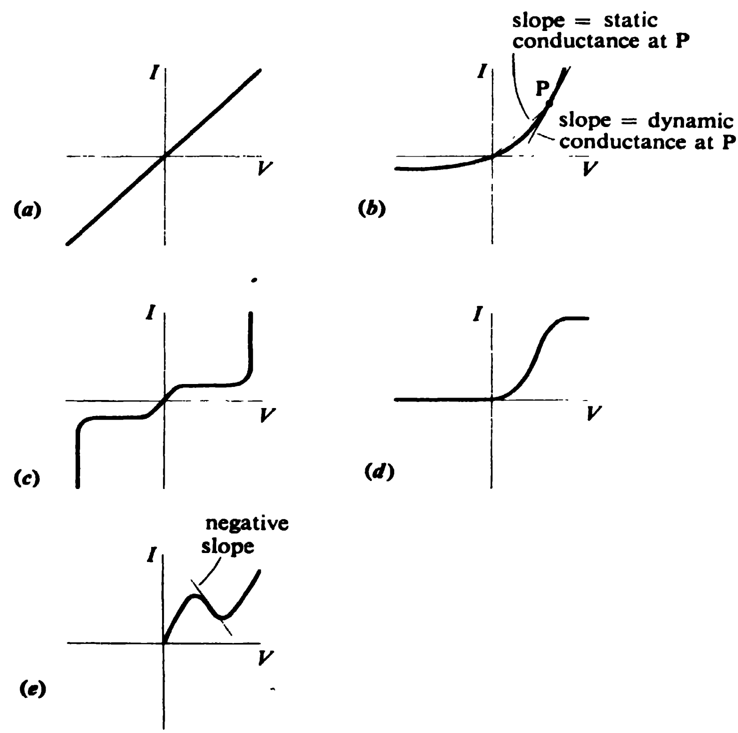


***Fig. 6.2.** Static characteristics of (a) a metal, (b) a typical rectifier, (c) a gas, (d) a thermionic diode, (e) a thermionic tetrode.*

with R and G whose properties can only be found by experiment. Conductors divide into two classes: *ohmic* or *linear* if R is constant at constant temperature, pressure, etc., and *non-linear* if R varies with the current. Most metals and many alloys are found to be linear, a property known as *Ohm's law.* Chrystal (1876) showed that conductors of iron, platinum and german silver with cross-sections of 1 cm^2 and resistances of 1 Ω for infinitesimally small currents did not change in resistance by as much as 1 part in 10^{12} when the current increased to 1 A, the temperature being held constant. Since then, there is ample evidence that little change occurs even when current densities of up to 10^5 A/cm^2 are used.

Graphs of steady current against steady potential difference are known as *static characteristics* for the conductor in question. Figure 6.2 illustrates some typical examples and shows that non-linear conductors may have linear parts in their characteristics so that the *dynamic* or *incremental resistance* dV/dI is constant over a small range of currents. The figures also show that incremental resistance and conductance may be negative where resistance itself never is.

Production of Heat. In section 3.7 we saw that a charge Q falling through a potential difference V loses potential energy QV. In a conductor carrying a steady current no kinetic energy is gained overall by the charges so that all the energy lost must be given to the conducting medium and appear eventually as heat. Thus, QV is the total heat produced in a conductor through which a total charge Q falls through V, and the rate of production of heat if the current is steady at I is IV. Thus

$$\text{Rate of production of heat in a conductor} = IV = RI^2 = GV^2 \tag{6.10}$$

in watts when R is in ohms and I in amps. For ohmic conductors, R is constant and Joule's law (1.2) is now seen as a consequence of our choice of the magnetic effect for the measurement of current, of the conservation of energy and of Ohm's law. Alternatively, we can look upon Joule's law and Ohm's law as jointly confirming that electric energy of charges (QV) must be included in the general law of conservation of energy.

Internal Resistance. The definition of resistance in (6.8) does *not* apply to conductors containing sources of e.m.f., for which a different approach is necessary. In supplying a current I, a source of e.m.f. $\mathscr{E}$ supplies energy at a rate $I\mathscr{E}$ and yet only IV appears as heat

in the external part of the circuit, V being the potential difference between the terminals. Hence $I(\mathscr{E}-V)$ must be the rate at which work is done by the source against its own internal resistance and must give the rate of production of heat within the source. By (6.10) it follows that the internal resistance must be given by this rate divided by I^2, that is, by $(\mathscr{E}-V)/I$. Hence the internal resistance r is given by

$$r = (\mathscr{E}-V)/I \quad \text{or} \quad \mathscr{E} = V+rI \quad \text{(Definition of } r) \qquad (6.11)$$

where I is the current in the *external* part of the circuit flowing *from* the positive terminal of the source *to* the negative. This is a very important relation and shows that $\mathscr{E} \neq V$ unless $r=0$ (unrealizable) or $I=0$ (the open-circuit condition (6.5)). If $\mathscr{E}$ is the only e.m.f., I is always positive and $\mathscr{E} > V$; but if I is negative, for instance in an accumulator being charged by another source of e.m.f., then $\mathscr{E} < V$.

Equation (6.11) means that the source behaves exactly like an e.m.f. $\mathscr{E}$ with zero internal resistance in series with a resistance r (Fig. 6.3) because both yield a potential difference $\mathscr{E}-rI$ when a current is supplied to any network connected across the terminals.

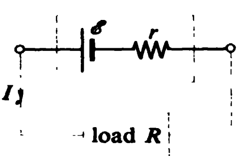


Fig. 6.3. Equivalent circuit for a source of e.m.f. with internal resistance.

Figure 6.3 is then said to be an *equivalent circuit* or network to the actual one.

6.3 Combinations of Resistors

Linear resistors connected by wires of negligible resistance form a *passive* or *inert* network and if there are n external connections they form an n-terminal network. We shall find that 2- and 4-terminal networks are the most frequently encountered.

A two-terminal passive network has an *equivalent resistance* defined by (6.8) where V is the potential difference between the terminals when a current I flows in at one and out of the other. The calculation of equivalent resistance follows a similar course to that of equivalent capacitance in chapter 5 and uses again the two

principles of conservation of charge and path-independence of potential difference. The first, in conjunction with $I = \mathrm{d}Q/\mathrm{d}t$ from chapter 1, means that, because the conditions are steady and no charge accumulates in any small portion of the wires, the total current entering a junction must equal that leaving (Fig. 6.4). Thus

$$\Sigma I = 0 \quad \text{at any junction} \quad \text{(steady currents)} \quad (6.12)$$

if a negative sign is allocated to current *entering*. This convention is adopted so that current *leaving* the shaded volume in Fig. 6.4

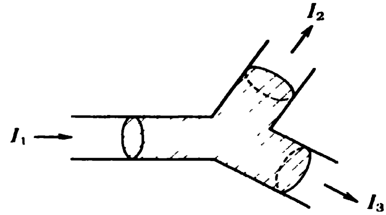


Fig. 6.4. ***Currents at a junction.*** $-I_1 + I_2 + I_3 = 0$ ***by the conservation of charge and*** $I = \mathrm{d}Q/\mathrm{d}t$.

shall be positive, in line with the sign allocated to outward normals in appendix 4.1. The path-independence of V is expressed by

$$\Sigma V = 0 \quad \text{round any closed conducting path} \quad (6.13)$$

Resistors in Series and Parallel. Rather than use (6.12) and (6.13) directly, it is often possible to divide a network into collections of series and parallel combinations. For resistors in series

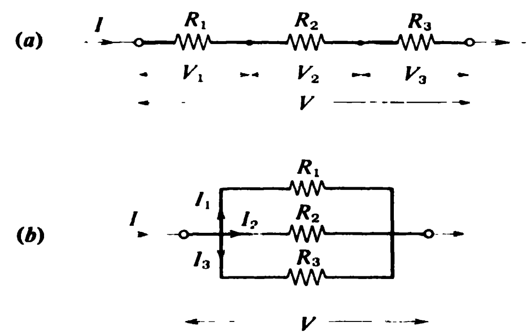


Fig. 6.5. ***(a) Resistances in series (common current); (b) resistances in parallel (common potential difference).***

(Fig. 6.5a), the current is common to all and the potential drops add to give a total potential difference $V = V_1 + V_2 + V_3 = I(R_1 + R_2 + R_3)$. Hence, in general,

$$R = \sum_i R_i \tag{6.14}$$

for resistors in series.

In parallel (Fig. 6.5b), the potential difference V is common to all while the currents add by (6.12), so that $I = I_1 + I_2 + I_3 = V(G_1 + G_2 + G_3)$ and hence

$$G = \sum_i G_i \quad \text{or} \quad \frac{1}{R} = \sum_i \frac{1}{R_i} \tag{6.15}$$

for resistors in parallel.

Delta-Y Transformation. This is useful in reducing networks which cannot be treated as collections of series and parallel combinations and is similar to that for condensers in problem 5.7. In Fig. 6.6a, by equating the equivalent resistances between A and B,

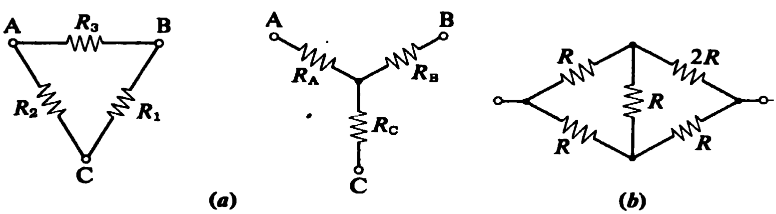


Fig. 6.6. (a) The delta-Y transformation; (b) a network which cannot be broken down into series and parallel combinations.

between B and C and between C and A, three relations are obtained which yield

$$\left.\begin{aligned} R_A &= R_2R_3/(R_1+R_2+R_3), \text{ etc.} \\ \text{and} \qquad R_1 &= (R_BR_C+R_CR_A+R_AR_B)/R_A, \text{ etc.} \end{aligned}\right\} \tag{6.16}$$

for the equivalence of the delta and Y.

The equivalent resistance of the network in Fig. 6.6b can be shown to be $13R/11$ either by using (6.16) or (6.12) and (6.13) as in the next section.

Just as any 2-terminal network can be replaced by an equivalent R, so a 3-terminal network can be replaced by a delta or a Y.

6.4 General Steady-Current Networks

A typical network containing sources of e.m.f. with internal resistances is shown in Fig. 6.7a, which also explains the terminology to be used. The problems encountered give the values of some e.m.f.s, resistances and possibly currents, and require others to be calculated. This can be done using (6.12) and (6.13), the latter modified as follows.

A quite general branch of a network (Fig. 6.7b) contains a source $\mathscr{E}$ with internal resistance r and other resistances totalling R'. The current I in this branch, counted positive when flowing through the source as if supplied by it, means that the potential difference across

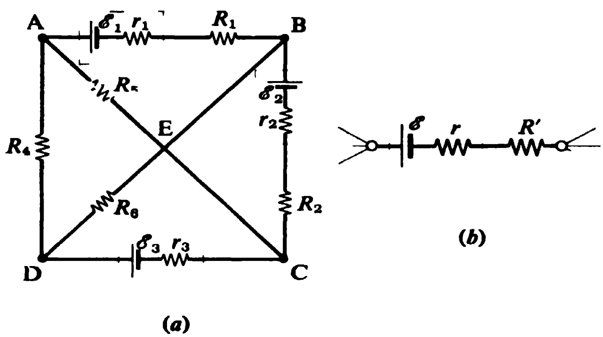


Fig. 6.7. (a) Illustrating a general network. Junctions *are indicated by dots as at* A, B, C, D; E *is not a junction.* AB, BC, *etc. are* branches. *Circuits such as* ABC, ABCD *are* meshes. *(b) A general branch.*

the branch is $\mathscr{E} - rI - R'I$ or $\mathscr{E} - RI$ where R is the total resistance in the branch. Round a mesh therefore, by (6.13), $\Sigma V = \Sigma (\mathscr{E} - RI) = 0$ or

$$\Sigma \mathscr{E} = \Sigma RI \tag{6.17}$$

the sums being algebraic and signs allocated to $\mathscr{E}$ and I by choosing positive senses round each mesh. Equations (6.12) and (6.17) are known as *Kirchhoff's first and second laws* respectively and are sufficient to solve any network problem.

Use of Kirchhoff's Laws. Slavish application of the laws is usually unwieldy and the following hints are given, using the network of Fig. 6.8 as an example in which all the currents are required.

{Avoid the use of subscripts in numerical problems: they are useful in theoretical work (cf. equation (6.19)) but the use, say, of α,

β, etc. for currents saves time and prevents confusion. Secondly, never write down explicit equations for unknown currents using Kirchhoff's first law, but reduce the number of unknowns by applying the law directly in a circuit diagram: thus if α is the current from A in Fig. 6.8 and β is the current from B then that in the 10 Ω

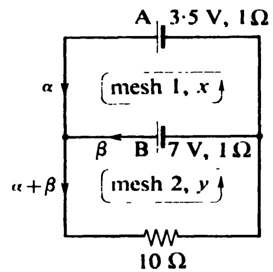


***Fig. 6.8.** Application of Kirchhoff's laws and circuital currents.*

resistor is $\alpha+\beta$ and a third unknown is avoided. Thirdly, apply the second law to as many *independent* meshes as there are unknowns: thus

$$3{\cdot}5-7 = \alpha-\beta \quad \text{in mesh 1} \tag{6.18}$$

$$7 = \beta+10(\alpha+\beta) \quad \text{in mesh 2}$$

from which α is $-1{\cdot}5$ A, β is 2 A and $(\alpha+\beta)$ is $-0{\cdot}5$ A.

Circuital Currents. Another method is to use circuital or mesh currents. For instance, in Fig. 6.8 let the current in mesh 1 be x and that in mesh 2 be y so that the current originally denoted by β is now $(y-x)$. This method ensures that Kirchhoff's first law is automatically satisfied, and by applying the second we obtain

$$3{\cdot}5-7 = x-(y-x) \quad \text{in mesh 1}$$

$$7 = (y-x)+10y \quad \text{in mesh 2,}$$

giving $x = -1{\cdot}5$ A and $y = -0{\cdot}5$ A as before.

Power Transfer. When sources of e.m.f. are used to operate power-consuming devices, any consumption in the internal resistance is wastage. In Fig. 6.3, the load of resistance R will carry a current $\mathscr{E}/(R+r)$, and will consume power $P = R\mathscr{E}^2/(R+r)^2$. For a given load R, P is a maximum for $r = 0$, but it is more common to find that r is fixed and that there is some choice of R. By equating

dP/dR to zero, the condition for maximum power transfer to the load is found to be $R=r$: the load is then said to be *matched* to the source.

Superposition Theorem. In a general network, let the meshes be labelled 1, 2, 3, etc. and let us use mesh currents I_1, I_2, etc. Then for each mesh Kirchhoff's second law gives

$$\left.\begin{aligned} \mathscr{E}_1 &= R_{11}I_1 - R_{12}I_2 - R_{13}I_3 - \cdots \\ \mathscr{E}_2 &= R_{22}I_2 - R_{12}I_1 - R_{23}I_3 - \cdots \text{etc.} \end{aligned}\right\} \qquad (6.19)$$

where $\mathscr{E}_1$ is the *total* e.m.f. in mesh 1, etc., R_{12} is the resistance common to meshes 1 and 2, etc., and R_{11} is the *total* resistance round mesh 1, etc. (see Fig. 6.9). Equations (6.19) can be solved for I_1, I_2, etc. to take the form

$$I_1 = A_{11}\mathscr{E}_1 + A_{12}\mathscr{E}_2 + A_{13}\mathscr{E}_3 + \cdots \qquad (6.20)$$

with corresponding equations for I_2, etc., in which the A's are functions only of the resistances. Equations (6.20) show that the

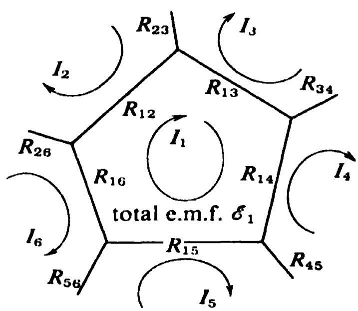


Fig. 6.9. A mesh in a general network. $R_{11} = R_{12} + R_{13} + \cdots + R_{16}$.

mesh currents, and hence the branch currents, superpose: put formally, if the current in a branch is I_1 due to an e.m.f. $\mathscr{E}_1$ acting alone at its proper place in the network and if the current in the same branch is I_2 due to $\mathscr{E}_2$ acting alone, the current in the branch when both $\mathscr{E}_1$ and $\mathscr{E}_2$ act is $I_1 + I_2$. It is understood that '$\mathscr{E}_1$ alone' means that all other e.m.f.s are replaced by inert resistances equal to their internal resistance.

Thévenin's Theorem. Consider the ith branch of a network between points A and B as in Fig. 6.10a in which there will in

general be an e.m.f. $\mathscr{E}_i$ and a total resistance R_i. Thévenin's theorem states that the current I in the branch may be obtained by replacing the rest of the network by an e.m.f. $\mathscr{E}_0$ with a series resistance R_0 as in Fig. 6.10c, where $\mathscr{E}_0$ is the potential difference which appears across AB if the ith branch is cut, and R_0 is the resistance of the rest of the network between A and B when all

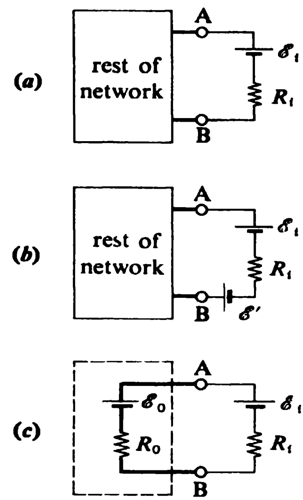


Fig. 6.10. Thévenin's theorem. (a) The actual situation; (b) derivation of the theorem by the addition of $\mathscr{E}$; (c) the equivalent circuit.

e.m.f.s in it have been replaced by inert resistances equal to their internal resistances. The proof is as follows:

Into the actual circuit of Fig. 6.10a insert an e.m.f. $\mathscr{E}'$ as in Fig. 6.10b so as to oppose $\mathscr{E}_i$, when the current in the ith branch will by superposition now be $I +$ the current due to $\mathscr{E}'$ alone. Hence

$$\text{New current} = I - \frac{\mathscr{E}'}{(R_i + R_0)} \tag{6.21}$$

If, however, $\mathscr{E}'$ is such that the new current is zero, then the combined e.m.f. $\mathscr{E}_i - \mathscr{E}'$ must equal the potential difference across AB when on open circuit—that is, must equal $\mathscr{E}_0$. Thus from (6.21)

$$0 = I - \frac{(\mathscr{E}_i - \mathscr{E}_0)}{(R_i + R_0)}$$

or

$$I = (\mathscr{E}_i - \mathscr{E}_0)/(R_i + R_0)$$

and this is just the current which would be produced by the equivalent circuit of Fig. 6.10c.

If the ith branch contains only an inert resistance R_i, then $I = \mathscr{E}_0/(R_i + R_0)$ and it is in this form that the theorem finds its most common applications.

6.5 The Charge and Discharge of a Condenser

While Kirchhoff's laws and the other results of sections 6.3 and 6.4 strictly refer to steady currents, we can assume that they apply at any instant for varying currents provided we check experimentally any deductions made.

An example is provided by the charging of a condenser C through a resistance R which in Fig. 6.11a includes all leads, internal resistance, etc. as well as any resistors. Let the current at a

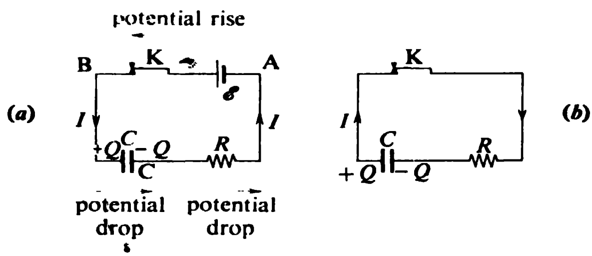


Fig. 6.11. (a) Charge and (b) discharge of C through R.

time t after the key K is closed be I in the direction shown and let the charge which has accumulated on C be $\pm Q$. Then, because the e.m.f. has no internal resistance, the rise in potential from A to B is $\mathscr{E}$, the fall from B to C is Q/C and the fall from C to A is RI. Thus, because $\Sigma V = 0$ round the circuit, $\mathscr{E} - Q/C - RI = 0$. But $I = \mathrm{d}Q/\mathrm{d}t$, so that

$$R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{Q}{C} = \mathscr{E} \tag{6.22}$$

Let $C\mathscr{E}$ be a new constant Q_0 so that (6.22) can be written

$$R\frac{\mathrm{d}(Q_0 - Q)}{\mathrm{d}t} + \frac{(Q_0 - Q)}{C} = 0$$

or

$$\frac{\mathrm{d}(Q_0 - Q)}{(Q_0 - Q)} = -\frac{\mathrm{d}t}{RC}$$

Integrating,

$$\log_e (Q_0 - Q) = -t/RC + \text{constant}$$

At $t=0$, $Q=0$ and the constant is thus $\log_e Q_0$. This finally gives

$$Q = Q_0(1 - e^{-t/RC}) \tag{6.23}$$

and
$$I = dQ/dt = \frac{Q_0}{CR} e^{-t/RC} \tag{6.24}$$

Q_0 is thus identified with the charge on the condenser after an infinite time, while the quantity RC, known as the *time constant* or the *relaxation time* of the circuit, gives the time for the charge to rise to $1-1/e$ or about $\frac{2}{3}$ of its final value. It thus gives an indication of the rate of charge as shown in Fig. 6.12a.

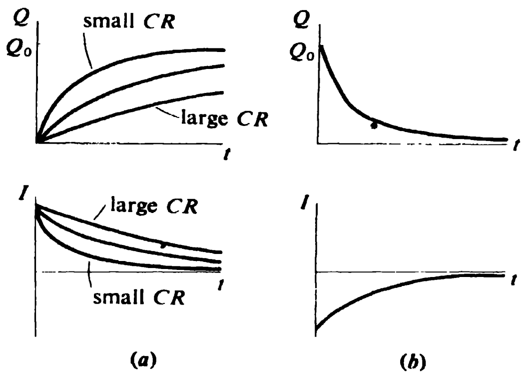


***Fig. 6.12.** Growth and decay of charge and current in a CR circuit. (a) Charging a condenser; (b) discharging a condenser.*

For the discharge of a condenser with an initial charge Q_0, let the current and charge after a time t be I and $\pm Q$ as in Fig. 6.11b, where this time $I = -dQ/dt$ (note the sign particularly). $\Sigma V = 0$ gives

$$R\,dQ/dt + Q/C = 0$$

yielding, by a similar method to that used above,

$$Q = Q_0 e^{-t/RC} \tag{6.25}$$

and
$$I = \frac{Q_0}{RC} e^{-t/RC} \tag{6.26}$$

plotted as negative in Fig. 6.12b because it flows in the opposite direction to (6.24) as Fig. 6.11 shows.

The charge remaining on the condenser after various times can be measured by a ballistic galvanometer (section 16.6) and a graph of $\log_e Q$ against t gives a straight line of slope $-1/RC$. In practice, the dielectric in the condenser will have a small conductivity producing an effective leakage resistance in parallel with C and R: this can be determined by carrying out the experiment without R.

For high resistances, the experiment can be used to determine R in terms of a standard condenser: for instance, with $C = 1\ \mu F$ and R about 50 MΩ the time constant is conveniently about 50 s.

Such experiments justify our assumption that steady current laws can be applied at any rate to slowly-varying currents, which are then described as *quasi-steady*.

6.6 Resistivity and Conductivity

The resistance of an ohmic conductor depends on its shape and size and on the direction of current through it as well as on the material of which it is composed. These factors can be separated for cylindrical conductors carrying current parallel to the axis because it follows from (6.14) and (6.15) respectively that the resistance of such conductors is proportional to length and inversely proportional to cross-sectional area (Fig. 6.13). Thus

$$R = \rho l/A \qquad \text{(Definition of } \rho) \qquad (6.27)$$

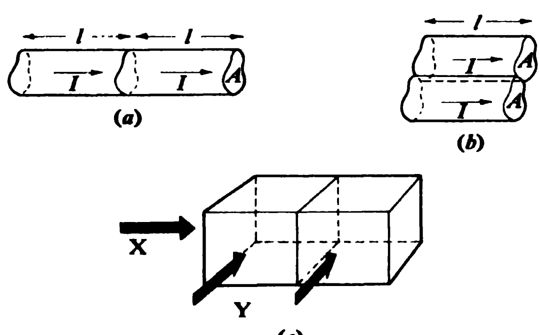


Fig. 6.13. Resistance of cylinders is (a) proportional to length, (b) inversely proportional to area of cross-section. The two cubes of (c) have twice the resistance of one for current in direction X, but only half for current in direction Y.

where ρ is a property of the material, its *resistivity*. The reciprocal is known as the conductivity σ, so that

$$\sigma = 1/\rho = l/RA \quad \text{(Definition of } \sigma\text{)} \quad (6.28)$$

The MKSA unit for ρ is the Ω-m and for σ the mho/m, although in practice the resistivity is more commonly quoted in Ω-cm. Table 6.1 gives the resistivities of some representative materials divided into three classes—good conductors ($\rho \sim 10^{-8}$ Ω-m), poor conductors and semiconductors ($\rho \sim 10^{-5} - 10^{4}$ Ω-m) and insulators ($\rho > 10^{10}$ Ω-m).

Table 6.1

CONDUCTIVITY AND RESISTIVITY OF SOME REPRESENTATIVE MATERIALS

	Resistivity (Ω-m) at 20°C	*Conductivity* (mho/m) *at* 20°C	*Mean temperature coefficient* (per °C) *between* 0°C *and* 100°C
Good conductors			
Silver	$1{\cdot}6 \times 10^{-8}$	$6{\cdot}2 \times 10^{7}$	0·0041
Copper	$1{\cdot}7 \times 10^{-8}$	$5{\cdot}8 \times 10^{7}$	0·0043
Aluminium	$2{\cdot}7 \times 10^{-8}$	$3{\cdot}7 \times 10^{7}$	0·0045
Platinum	$10{\cdot}6 \times 10^{-8}$	$0{\cdot}94 \times 10^{7}$	0·0039
Manganin	$42{\cdot}0 \times 10^{-8}$	$0{\cdot}24 \times 10^{7}$	~0·00001
Constantan	$48{\cdot}0 \times 10^{-8}$	$0{\cdot}21 \times 10^{7}$	~0·00002
Mercury	$96{\cdot}0 \times 10^{-8}$	$0{\cdot}10 \times 10^{7}$	0·0010
Carbon (graphite)	$350 - 6{,}300 \times 10^{-8}$	$0{\cdot}16 - 2{\cdot}9 \times 10^{5}$	−0·0005
Poor conductors and semiconductors (denoted by S)			
InSb (S)	$5{\cdot}7 \times 10^{-5}$	$1{\cdot}76 \times 10^{4}$	
Saturated NaCl solution	$4{\cdot}4 \times 10^{-2}$	22·6	−0·02
Germanium (S)	0·47	2·1	
Silicon (S)	$2{\cdot}3 \times 10^{3}$	$4{\cdot}3 \times 10^{-4}$	
Distilled water	$5{\cdot}0 \times 10^{3}$	$2{\cdot}0 \times 10^{-4}$	
Insulators (see also table 13.1)			
Pyrex glass	$\sim 10^{12}$	$\sim 10^{\;12}$	
Paraffin wax	$\sim 10^{14}$	$\sim 10^{-14}$	
Polystyrene	$\sim 10^{15}$	$\sim 10^{-15}$	

Data for good conductors and insulators are taken from Kaye and Laby (1959), and for semiconductors from Smith (1919), with the publishers' permission.

The statement often encountered that the resistivity of a material is the resistance per unit cube is incorrect: it implies that the unit of ρ is the Ω/m^3 which is clearly not in accord with (6.27), and it ignores any reference to the current flow. Even if the direction of this flow is specified as parallel to one pair of sides, the definition still

implies that two unit cubes have a resistance twice that of one, whereas it may be halved (Fig. 6.13c). The correct definition (6.27) contains the understanding that l is a dimension parallel to the current and A an area perpendicular to it.

By dividing a conductor into small cylinders each of resistance $l/\sigma A$ it is often possible to obtain a total resistance by integration. For instance, if a current flows radially outwards between two concentric spherical surfaces of radii a and b as in Fig. 6.14, then a typical shell has a resistance $\mathrm{d}r/4\pi r^2\sigma$ (elements of the shell in parallel). All the shells are in series so that the total resistance is the integral of $\mathrm{d}r/4\pi r^2\sigma$ between a and b, i.e. $(b-a)/4\pi\sigma ab$.

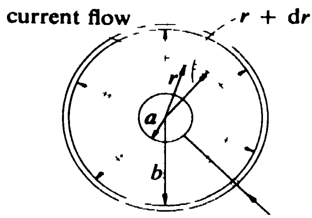


Fig. 6.14. Radial current flow in a conductor between concentric spherical surfaces.

Distributed Resistance. While it is often possible to neglect the resistance of leads to resistors or to other apparatus supplied from a source of e.m.f., the transmission of power to a load sometimes takes place through cables so long that the potential drop is serious unless the resistance per unit length or the current are kept small. For overhead cables in the national grid, the diameter cannot be increased indefinitely because of weight and cost and the loss is reduced by using high voltages; for distribution within a building weight is not a limitation and larger diameters are used in preference to higher voltages for obvious reasons.

The potential drop within a uniform wire is itself uniform, and Fig. 6.15 shows equipotentials, lines of force and a potential diagram for a wire carrying a steady current connected to a source of e.m.f. A series circuit contains just such a set of potential drops (cf. section 6.5).

Variation of Resistivity. The resistivity of both linear and nonlinear materials may vary when physical conditions are changed. The resultant change in resistance can be obtained by simple and accurate experiments (section 6.8) and is thus often used to measure

the quantity producing the change. As examples we consider temperature, magnetic field and mechanical strain.

The resistance of metals and alloys increases with rise in *temperature* according to the empirical law

$$R_{\theta_1} = R_\theta[1+\alpha(\theta_1-\theta)+\beta(\theta_1-\theta)^2+\cdots] \tag{6.29}$$

where θ and θ_1 are temperatures, and α, β, etc. are coefficients decreasing rapidly in magnitude from one to the next. Typical values are 0·004/°C for α and 10^{-6}/°C^2 for β so that only α, the

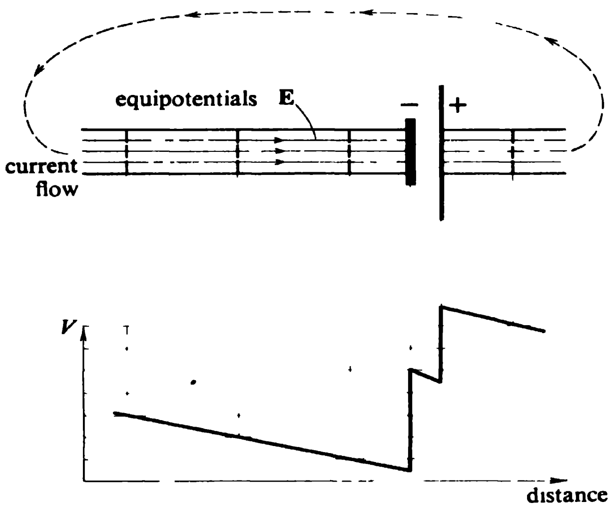


Fig. 6.15. Equipotentials, lines of force and potential diagram for a steady current flowing in a uniform wire.

temperature coefficient of resistance, need normally be considered: it is listed for some materials in table 6.1. Because α has a value close to 1/273 for many metals, a temperature scale furnished by the variation of resistance is very approximately the same as that on the ideal gas scale and platinum resistance thermometers can be corrected to such a scale if α and β for Pt are known (see Roberts and Miller, 1921). Equation (6.29) also gives the variation of *resistivity* with temperature although the values of α, β, etc. are slightly different because of the changes in dimensions: the difference for moderate $(\theta_1-\theta)$ is negligible.

(At very low temperatures, usually below 10°K, the resistance of many metals and alloys suddenly decreases to a negligible value, a phenomenon known as *superconductivity*. Semiconductors, insulators and electrolytes all have resistivities which decrease as temperature rises: this and other aspects of conduction are discussed in chapter 12.

Bismuth shows a remarkably high variation of resistivity when a *magnetic field* is applied at right angles to the current flow and this property has been used to measure the strength of such fields: other materials show extremely small changes by comparison. The phenomenon is known as *magnetoresistance* and is allied to the Hall effect (section 12.4).

A stretched metal wire undergoes a measurable change in resistance which can be used as a measure of *mechanical strain*. A fine wire bent backwards and forwards until it covers an area of about the size of a postage stamp is sandwiched between two fine pieces of paper and the whole strain gauge thus formed is made to adhere to the surface whose strain is required. The gauge is strained by the same amount as the surface and the change in resistance is measured.

6.7 Resistors in Practice

Fixed resistors fall broadly into three groups: those of high stability and known to at least fair accuracy for use in measurements; power resistors capable of dissipating up to several hundred watts; and general purpose resistors whose values need not be known with precision, which dissipate at most a few watts and which are chiefly for use in electronic circuits.

For the first group, the material usually used is manganin wire or sheet which combines a relatively large resistivity with a very small temperature coefficient of resistance and a small thermal e.m.f. against copper. Where the last property is unimportant, constantan and some nickel chrome alloys are preferable because of their resistance to corrosion. For A.C. work, other factors must be considered (section 10.12).

Power resistors are wire-wound, and general purpose resistors may be 'composition' (graphite, filler and resin baked into a rod), 'cracked-carbon' (carbon deposited at a high temperature on to a ceramic rod) or 'metal-film' in increasing order of stability: carbon resistors are cheap and compact but rarely used in measuring circuits.

Variable resistors are either (*a*) resistance boxes giving resistances in steps by removable plugs or by dials controlling a decade as in Fig. 6.16 or (*b*) rheostats with sliding contacts. The latter can also

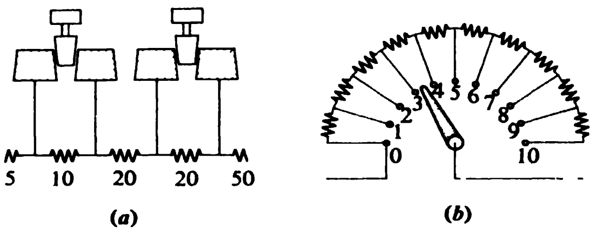


Fig. 6.16. Arrangement of coils in (a) a section of a plug type resistance box, (b) one decade of a dial type resistance box.

be used to obtain variable potential differences from a fixed source as shown in Fig. 6.17b and it is then known as a potential divider: electronic engineers often refer to it mistakenly as a potentiometer,

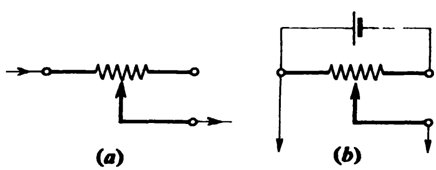


Fig. 6.17 A rheostat used as (a) a variable resistance, (b) a potential divider.

a term we shall reserve for the measuring instrument described in section 6.8.

In all cases care must be taken that the resistor can dissipate the required power.

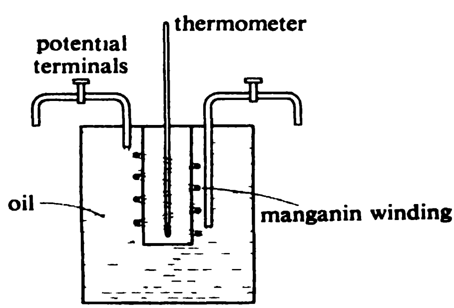


Fig. 6.18. Construction of a standard resistor.

Standard Resistors. These are constructed of manganin wire sealed into a container often filled with oil whose temperature can be measured and they are fitted with thick copper terminals as in Fig. 6.18. For standards of less than about 1 Ω, the use of only two terminals makes the resistance indefinite, not only because of leads (whose resistance can be allowed for) but principally because of *contact resistances* at the terminals themselves. These are due to contamination of the surface and can amount to 0·0001 Ω even when care is taken: this is a serious error particularly with low resistances and it is, moreover, variable. Consequently, four terminals are provided as in Fig. 6.19, the resistance R being defined

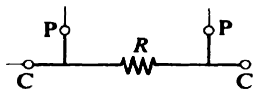


Fig. 6.19. Current and potential terminals.

as the ratio of the potential difference between PP, the potential terminals, to the current between CC, the current terminals. It is clear that contact resistances at CC do not affect R and those at PP are arranged to produce zero or negligible effect by the circuit connections (see next section).

6.8 D.C.* Measurements

A laboratory is normally equipped with standard resistors of various values and with standard Weston cells (section 12.8) in terms of which nearly all D.C. measurements can be made. We consider here only the main principles of such measurements: for more details the references at the end of the chapter should be consulted.

Measurement of Potential Difference. The potentiometer, one of the most important of electrical instruments, consists basically of a resistance through which a steady current flows as shown in Fig. 6.20a: when a potential difference V is connected across R_1 and the tapping adjusted until no current flows in the galvanometer G, application of $\Sigma\, V=0$ to the circuit ABV gives $V=R_1 I$. When V is replaced by a standard cell of e.m.f. $\mathscr{E}_s$ the balance-point shifts to

* The term *direct current* or D.C. will here be used to mean a steady current although it is occasionally used for any unidirectional current.

give a resistance R_s so that $\mathscr{E}_s = R_s I$ and

$$V = \frac{R_1}{R_s}\mathscr{E}_s \tag{6.30}$$

provided I has remained constant. Thus, although the driving e.m.f. $\mathscr{E}$ and the resistances r and R do not need to be known, they must be constant from the time the first balance-point is obtained:

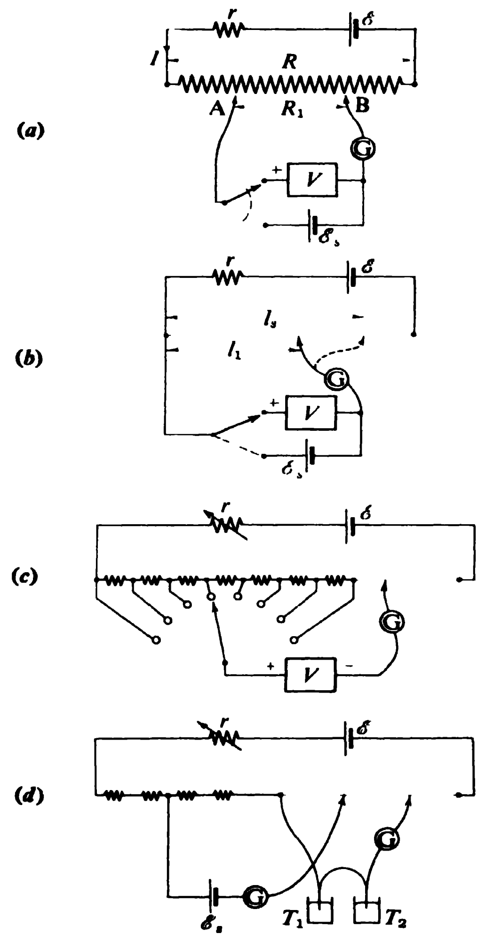


Fig. 6.20. Development of the potentiometer.

it is sometimes convenient to attain this first balance-point by fixing the tapping point B and adjusting I by means of r, as in the achievement of a direct-reading instrument below. A balance-point will only be found on R if the potential differences are connected to A and B with the same polarity and if they are smaller than that across R.

The particular virtue of the method lies in the absence of current drawn from the source of potential difference: the potentiometer thereby acts like a voltmeter of infinite resistance and can measure e.m.f.s. A further consequence, common to all null methods, is that the calibration of the galvanometer is of no importance.

Some precision potentiometers have R consisting of sets of high quality resistance coils, but most combine such coils with a slide-wire on which the actual balance point is obtained as in Figs. 6.20b and c. The simple circuit of Fig. 6.20b depends upon the uniformity of the wire to make R_1/R_s of (6.30) equal to l_1/l_s, while in Fig. 6.20c the coils in series with the slide-wire are each of the same resistance as the wire and thus form an extension of it. This is particularly valuable for small potential differences such as are produced by thermocouples (Fig. 6.20d). For large potential differences and e.m.f.s a potential divider must be used as in Fig. 6.21: the potential

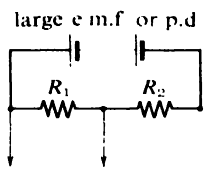


Fig. 6.21. Measurement of large potential difference.

difference measured is only $R_1/(R_1+R_2)$ of that required but there is bound to be some error because current is taken from the source.

When the resistance r of Fig. 6.20c is adjusted so that a standard cell gives a balance point across coils and wire of known length it is possible to recalibrate the instrument directly in V or mV etc. and thus make it direct-reading (see problem 6.22).

Voltmeters of various kinds are described elsewhere, but they can all be calibrated by measuring various potential differences simultaneously with the meter and a potentiometer itself calibrated in terms of a standard cell. Voltmeters provide a convenient secondary method for measuring potential difference but are inaccurate

unless their resistance is much higher than that across which they are connected (as in electrometers, section 5.8).

Measurement of Resistance. For resistances from about 1 Ω to 1 MΩ and possibly higher, depending on the accuracy required, the Wheatstone bridge is the basic network for comparing an unknown with a standard resistance. The condition for zero current through the galvanometer in Fig. 6.22 is that B and D should be at the same potential so that $P\alpha = R\beta$ and $Q\alpha = X\beta$. Hence $P/Q = R/X$ independent of the e.m.f. $\mathscr{E}$. If the unknown is X, the absolute values

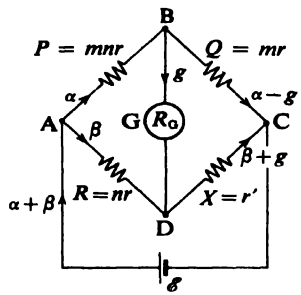


Fig. 6.22. The basic Wheatstone bridge.

of P and Q are not important but only their ratio, and this can be obtained either by using a uniform slide-wire so that the ratio is equal to that of the lengths of wire (the so-called metre bridge) or by using sets of resistors whose ratios are accurately known. The only standard needed is R.

An important property of the bridge is its sensitivity S, defined as the deflection of the galvanometer, $d\theta$, for a small fractional out-of-balance resistance dr/r in one arm, using the second notation of Fig. 6.22. Thus $S = r\, d\theta/dr$. The four arms with resistances mnr, mr, nr and r' form a general bridge for which the balance condition will be $r = r'$ (Callendar, 1910). Applying Kirchhoff's second law to the meshes ABD, BCD:

$$mnr\alpha - nr\beta = -R_G g = r'(\beta + g) - mr(\alpha - g)$$

Eliminating α gives

$$g = \frac{\beta(r - r')}{(mr + r') + R_G(1 + 1/n)}$$

For a bridge just out of balance, $(r-r')\rightarrow \mathrm{d}r$ and $g\rightarrow \mathrm{d}g$ so that

$$\mathrm{d}g = \frac{\beta\,\mathrm{d}r}{r(m+1)+R_G(1+1/n)} \tag{6.31}$$

and, because β is very nearly $\mathscr{E}/(n+1)r$, neglecting internal resistance,

$$\mathrm{d}g = \frac{\mathscr{E}\,\mathrm{d}r}{r(n+1)[(m+1)r+R_G(1+1/n)]}$$

The sensitivity of the bridge is thus

$$S = r\frac{\mathrm{d}\theta}{\mathrm{d}r} = r\frac{\mathrm{d}\theta}{\mathrm{d}g}\frac{\mathrm{d}g}{\mathrm{d}r} = \frac{\mathscr{E}S_I}{(n+1)[(m+1)r+R_G(1+1/n)]} \tag{6.32}$$

where S_I is the current sensitivity of the galvanometer. To obtain a high value for S, m should be as small as possible and n should not have an extreme value, but in practice as long as m is not large and n is approximately unity, their values have little effect on S. $\mathscr{E}$ cannot be increased indefinitely because of heating effects in the arms. S_I depends largely on $R_G^{1/2}$ (section 16.5) and this gives an optimum value for R_G equal to the resistance of the bridge as viewed from the galvanometer terminals—R_G is matched to the bridge: this is the most important practical condition.

The bridge is not suitable for high resistances because S falls as r increases, nor for low resistances because there is no provision for eliminating lead or contact resistances.

The Carey-Foster bridge (Fig. 6.23) is useful for resistances of about 1 Ω because lead resistances are eliminated and non-uniformity of the slide-wire can be measured and allowed for. If $R - X$

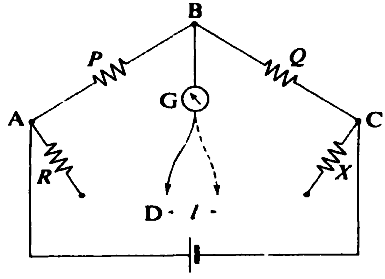


Fig. 6.23. The Carey-Foster bridge.

is x and a balance-point is obtained as shown, then the interchange of R and X *without leads* effectively transfers x from arm AD to arm DC. To restore balance, the point D must shift to the right by a length of wire whose resistance is x. If the resistance per unit length of the wire is r

$$x = R - X = rl \tag{6.33}$$

where l is the shift of the balance-point. P and Q do not affect the result but a change in their ratio does shift the *two* balance-points to a different part of the wire. By using a known R and X in (6.33), r can be found for any length of wire, and thereafter an unknown X can be found in terms of a standard R. If the wire is non-uniform, a known but small $(R - X)$ can be used to obtain r for a short length of the wire, and the whole wire calibrated by varying P.

Low Resistances. Since low resistances are fitted with current and potential terminals, any methods for their determination must ensure that they are used properly. Down to about 0·001 Ω (and exceptionally to 10^{-5} Ω), the Kelvin double bridge (Fig. 6.24) ensures that the contact and lead resistances are either in series with much higher resistances or are in parts of the network where the

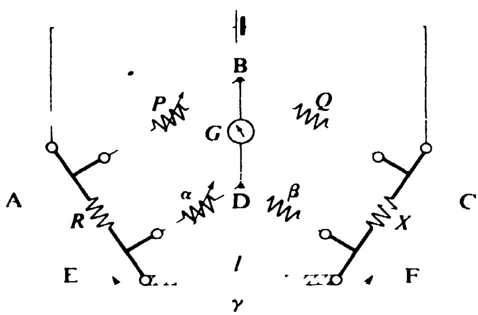


Fig. 6.24. The Kelvin double bridge.

value is immaterial. P, Q, α and β are resistors of normal magnitude, R and X are the standard and unknown low resistors. In one common method, P, Q, α and β are included in a double ratio arm box and P and α are varied together so that the ratios P/Q and α/β are always equal:

$$\frac{P}{Q} = \frac{\alpha}{\beta} \tag{6.34}$$

When the bridge is balanced by the simultaneous adjustment of P and α, the condition is obtained by transforming the delta DEF to a Y giving

$$\frac{P}{Q}=\frac{R+\dfrac{\alpha\gamma}{\alpha+\beta+\gamma}}{X+\dfrac{\beta\gamma}{\alpha+\beta+\gamma}} \tag{6.35}$$

where γ is the total resistance between E and F. Equating the right-hand sides of (6.34) and (6.35) and cross-multiplying gives $R/X=\alpha/\beta$ and from (6.34) both are equal to P/Q: α and β need not be known. In an alternative method of using the bridge, l is a thick link of copper which can be removed, while the condition (6.34) is not imposed by a ratio box. If P and α are used to obtain balance with l in and then out, (6.35) is the balance condition with l in, and $(R+\alpha)/(X+\beta)=P/Q$ with l out. This again gives $P/Q=\alpha/\beta=R/X$.

For resistances lower than about 0·001 Ω, a potentiometer method is often used as shown in Fig. 6.25. Here the contact resistances in

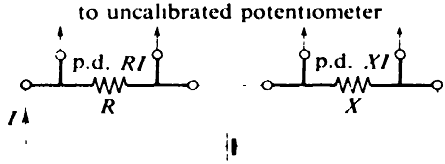


Fig. 6.25. Measurement of low resistance using a potentiometer.

the potential terminals carry no current at balance and thus do not affect the potential differences.

High Resistances. For resistances of several megohms and upwards, the condenser discharge method described in section 6.5 can be used. Alternatively, if a standard resistance of the same order of magnitude is available, a simple substitution method is possible (Fig. 6.26) because R and X are so much larger than any other resistance: the currents are proportional to $1/R$ and $1/X$ respectively and the deflections θ_R and θ_X of a linear galvanometer will thus be in the ratio X/R.

Resistance can also be measured quickly but relatively inaccurately using a calibrated voltmeter and ammeter (see problem 6.21).

Measurement of Current. Accurate measurement of current involves passing it through a standard resistance R and measuring the potential difference V across R by a potentiometer calibrated in terms of a standard cell. Then $I = V/R$. This method can be used

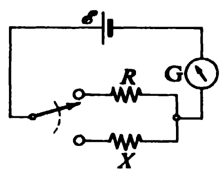


Fig. 6.26. The substitution method for measuring high resistances.

to calibrate an ammeter by connecting it in series with R. For details of electromagnetic current meters see sections 16.5 and 16.6.

Thermal E.m.f.s. These may be troublesome in all D.C. measurements, but they and their elimination are more appropriately discussed in section 12.7.

6.9 The Carriers of Charge in Conductors

From sections 1.5 and 3.9 we see that the carriers of charge can be identified by their specific charge (e/m), but we have not so far considered any evidence relating to the charge carriers in conductors. This is largely because it is completely immaterial as far as any phenomena so far encountered are concerned. Now, however, we wish to look at such evidence.

Electrons ($e/m = -1{\cdot}76 \times 10^{11}$ C/kg) are known to be emitted from the surfaces of many metals when at high temperatures (*thermionic emission*) and when irradiated with visible or ultraviolet light (*photo-electric emission*): there is thus a strong presumption that electrons are not tightly bound to metals and are responsible for electric conduction in them.

Electromechanical Experiments. Two types of experiment have been carried out with the aim of showing inertia effects due to the motion of charges in metals carrying a current. In the first, a conductor with no current is given a sudden acceleration or deceleration which should result in a measurable pulse of charge. Tolman and Stewart (1917) and their co-workers have conducted numerous experiments of this type and have been able to show that the carriers in copper have e/m about $-1{\cdot}93 \times 10^{11}$ C/kg, with similar values for silver and aluminium. These experiments are very

difficult to perform with great accuracy and the converse type has yielded more convincing results. This second method changes the current suddenly in a coil and measures the resultant change in motion. Barnett was the first to adopt the method in 1921, but that of Kettering and Scott (1922) will be briefly described.

If a coil of radius a (Fig. 6.27) carries a current I consisting of n

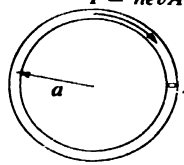


Fig. 6.27. The circular coil of Kettering and Scott's experiment.

charges per unit volume each of charge e, mass m and with a mean angular velocity ω, then by (1.15)

$$I = nea\omega A \tag{6.36}$$

where A is the cross-sectional area of the wire. The angular momentum of the charges is

$$L = \Sigma\, ma^2\omega = n2\pi aAma^2\omega$$

or

$$L = I\, 2\pi a^2(m/e)$$

using (6.36). Thus if a sudden change $\delta I = 2I$ occurred in the current by its reversal, the change in angular momentum would be

$$\delta L = 2I\, 2\pi a^2(m/e) \tag{6.37}$$

The coil is suspended in a horizontal plane as a torsional pendulum whose undamped oscillations have a period T and an amplitude θ. If the angular velocity at the centre of the oscillations is Ω, then by the conservation of energy $\frac{1}{2}c\theta^2 = \frac{1}{2}I_m\Omega^2$ where c is the torsional constant of the suspension and I_m is the moment of inertia of the coil. The angular momentum at the centre of the oscillations is thus

$$L_0 = I_m\Omega = I_m\theta(c/I_m)^{1/2} = 2\pi I_m\theta/T \tag{6.38}$$

since $T = 2\pi(I_m/c)^{1/2}$.

Equation (6.38) is a relation between L_0 and θ, so that if the reversal of current takes place at the centre point of the oscillations,

the change given by (6.37) occurs in L_0 and this gives a corresponding change in θ:

$$\delta\theta = T\,\delta L_0 / 2\pi I_{\mathrm{m}} = \frac{2Ia^2T}{I_{\mathrm{m}}}\frac{m}{e}$$

and all quantities but m/e can be measured.

Kettering and Scott found that the carriers in copper and aluminium were negative with $e/m = 1{\cdot}77 \times 10^{11}$ C/kg and could therefore be identified as electrons. Brown and Barnett (1922) carried out similar experiments on molybdenum and zinc and Scott (1921) on cadmium, all with similar results.

By contrast with metals, inertia experiments with electrolytes by Tolman (1921) confirm that the charges there responsible for the carriage of current are the positive and negative *ions* present in solution.

We shall leave further consideration of conduction until chapter 12.

6.10 Summary of Chapter 6

The important relations of this chapter result from definitions, first of e.m.f. by

$$\mathscr{E} = \oint \mathbf{E}\cdot d\mathbf{s} = \oint \mathbf{E}_{\mathrm{M}}\cdot d\mathbf{s} \qquad (6.2) \text{ and } (6.3)$$

and secondly of resistance, by

$$R = V/I \qquad (6.8)$$

Most of our concern has been with ohmic or linear conductors, for which R is independent of I or V: the linearity of the equations for steady-current networks is a consequence of this and has enabled us to derive some theorems (Kirchhoff's laws, superposition, Thévenin) useful in solving network problems. We have also seen that these theorems still seem to apply to linear networks when currents vary slowly in time.

References

Resistors: Dummer (1959). D.C. Measurements: Harris (1952); Stout (1960); Buckingham and Price (1955).

PROBLEMS

SECTION 6.1

6.1 A 45 ampere-hour battery will supply 45 A for 1 hr, 2 A for 22·5 hr, etc. If such a battery has an e.m.f. of 12 V, find (*a*) the total charge which can be supplied, (*b*) the total energy available, (*c*) how long the sidelights of a car could be operated if they consume 24 W, (*d*) how long the battery could operate the starter motor taking 135 A. Assume that the e.m.f. remains constant until complete discharge.

SECTION 6.2

6.2 A D.C. source has an e.m.f. of 120 V and a negligible internal resistance. If n cells each of e.m.f. 2·1 V and internal resistance 0·01 Ω are to be charged from this source with a charging current of 3 A, find the series resistance necessary. If $n = 20$, what proportion of energy delivered by the source is wasted as heat? Find also the potential difference across the 20 cells.

SECTION 6.3

6.3 Show that a Wheatstone bridge network with all resistances, including that of the galvanometer, equal to R presents a resistance R to the driving battery.

6.4 A skeleton cube is made of wires soldered together at the cube corners, the resistance of each wire being R. Calculate the equivalent resistance of the network between (*a*) diagonally opposite corners, (*b*) the two ends of one wire. Solve this classic problem either by sending a current I into the network and using Kirchhoff's laws and symmetry to find the currents in the wires or by using the Y-delta transformation.

6.5 In the ladder network of Fig. 6.28 all the resistances are equal to R. Find the resistance between the input terminals when the network is completed up to AA, up to BB, etc. What is the input resistance if the network is of infinite length?

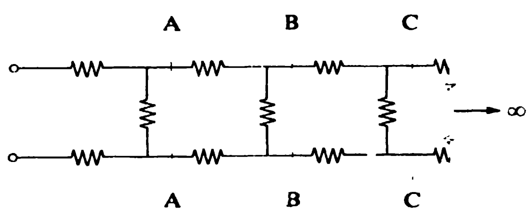


Fig. 6.28. Ladder network for problem 6.5.

6.6 Find the resistance presented to the battery in a network like that of the Kelvin double bridge of Fig. 6.24 if the resistances P, G, β, α, R, X and γ are all 1 Ω, and if Q is zero. Show that the same resistance is presented if β is 8 Ω instead of 1 Ω and explain why this is so.

*6.7 Eight resistors are of identical appearance but seven are of 1 Ω while the eighth is of 2 Ω. Can a 2-terminal network be constructed out of these resistors of such a form that *one* measurement of resistance across it would enable the odd resistor to be unambiguously selected?

6.8 In a simple circuit containing only a steady source of e.m.f. and resistances in series, a 100 Ω rheostat varies the current between 0·4 A and 2 A when its resistance is changed from 100 Ω to zero. Find the e.m.f. and the resistance in the rest of the circuit. The rheostat is found to be too coarse a control for a particular purpose: explain how to use in addition (*a*) a 10 Ω rheostat and (*b*) a 1 kΩ rheostat to achieve finer control.

SECTION 6.4

6.9 In a Wheatstone bridge network, the galvanometer is replaced by a battery of the same e.m.f. $\mathscr{E}$ as the one already present. If the resistances of the arms and of the batteries are all equal to R, find the currents in the various branches.

6.10 A D.C. generator in series with a resistance R is connected in parallel with both a battery of e.m.f. 12 V and a resistive load. The e.m.f. of the generator may fluctuate, and R is of such a value that all the current to the load is supplied by the generator when its e.m.f. is 60 V. What fraction of the current does it supply when its e.m.f. drops to 50 V? Neglect the internal resistance of the generator and battery.

6.11 Solve the problem of Fig. 6.8 by using the superposition theorem.

6.12 Three of the four arms of a Wheatstone bridge network are of 40 Ω, the fourth is 41 Ω and the resistance of the galvanometer is 20 Ω. If the cell has an e.m.f. of 2 V and negligible internal resistance, find the current through the galvanometer using (*a*) Kirchhoff's laws (*b*) Thévenin's theorem.

*6.13 How much of the network theory developed in this chapter depends on the linearity of the resistances? If Ohm's law does not apply but resistance is still defined as V/I, which of the formulae, etc. will still apply?

6.14 A current I divides between two resistances R_1 and R_2 in parallel. Show that the current through each resistance is correctly given by assuming that the total rate of production of heat is a minimum together with Kirchhoff's first law.

SECTION 6.5

6.15 A condenser of capacitance C_1 has a charge Q_0 on its plates. It is connected at time $t=0$ in series with a condenser of capacitance C_2 and a resistance R by the closing of a key. Find the time constant of the ensuing discharge.

SECTION 6.6

6.16 Calculate the resistance to a radial current between two concentric spherical surfaces of radii a and b when a material of conductivity σ fills the space between them. Is there anything familiar about the result?

6.17 A copper wire of radius a forms the inner conductor of a coaxial cable and an imperfect insulator of conductivity σ and external radius b fills the space between the inner and outer conductors. Find the resistance per unit length of the insulator for currents flowing in it parallel to the wire, and the conductance per unit length for currents flowing radially outwards from the wire. Why is the conductance rather than the resistance asked for in the second case?

6.18 A parallel-plate condenser has an imperfect insulator of conductivity σ and relative permittivity ε_r between its plates. What is the time constant of the self-discharge of the condenser?

6.19 A twin cable carries current between A and B, 5 miles apart. It is known to have one fault, i.e. a breakdown in the insulation between the two conductors (but not necessarily a short-circuit). A potential difference of 200 V maintained across the ends at A produces a potential difference of 50 V across the ends at B, while 200 V across B produces 20 V across A. Locate the fault.

SECTION 6.7

6.20 In the potential divider network of Fig. 6.17b, let the total resistance of AB be R, the resistance of AC be R_1 and of the load R_L. Find the input resistance (resistance as seen from the input terminals) and its value for large and small loads.

SECTION 6.8

6.21 A quick method of measuring a resistance R uses a voltmeter (resistance R_V) and an ammeter (resistance R_A) to measure the potential difference and current. There are two ways of connecting the meters (voltmeter across R only, voltmeter across R and the ammeter) both of which yield inaccurate values for R. If V is the voltmeter *reading* and I the ammeter *reading*, find expressions for the true value of R in terms of the apparent value, V/I, and deduce under what circumstances one method of connection should be used rather than the other.

6.22 How can a metre potentiometer wire of resistance 2 Ω be converted into a direct reading instrument in which 1 cm of wire corresponds to a potential difference of 1 mV? (E.m.f. of standard cell = 1·0183 V; approximate e.m.f. of driving cell = 2 V.)

6.23 In a simple potentiometer network, the total resistance round the driving cell circuit is R and the e.m.f. of the driving cell is $\mathscr{E}_1$. A second cell of e.m.f. $\mathscr{E}_2$ has its positive terminal connected to the same end A of the potentiometer wire as that of $\mathscr{E}_1$ and is further connected to a galvanometer and jockey in the conventional way. The cell $\mathscr{E}_2$ and galvanometer have a combined resistance r. If the jockey makes contact at X and if the resistance of AX is R_1, find the current in the galvanometer and the condition for it to be zero.

CHAPTER 7

MAGNETIC FIELDS AND MAGNETIC DIPOLES

In chapters 2 to 5 the *electric* forces between static charges *in vacuo* were thoroughly examined: we must now investigate in similar detail *magnetic* forces first mentioned in section 1.2. There, however, the existence of so-called permanent magnets was ignored so we begin in section 7.1 by surveying elementary phenomena connected with magnets and their interaction with currents. We shall also justify the use of the adjective 'magnetic' to describe magnet–magnet, magnet–current and current–current forces.

We need next a basic entity in terms of which fundamental magnetic laws may be expressed, just as Coulomb's law needed the concept of the point charge. Traditionally, the *magnetic pole* has been used for permanent magnets and the *current element* for steady currents. We shall eventually deal with these, but as a means of introducing the laws of magnetism both have their defects and we shall adopt the *magnetic dipole* as a basis, partly because it stresses from the start the unity of magnetism, whether arising from currents or magnets, and partly because, unlike the other two entities, we can obtain actual dipoles and experiment with them.

Such experiments are outlined in section 7.2 and lead to the establishment of a magnetic dipole analogous to the electric dipole, but because the basic law is so complex we immediately introduce the intermediate concept of magnetic induction or flux density. This enables us to express the magnetic law simply in two parts just as **E** (section 3.10) enabled Coulomb's law to be similarly expressed.

The rest of the chapter is devoted to examining the extent of the analogy between permanent magnetism and electrostatics.

7.1 Magnetic Forces and Couples

Magnet and Magnet. Permanent magnets are recognized by two effects known to the ancient Greeks, the *directive* and the *attractive*. The first effect is that, when permanent magnets are freely suspended

so that they can oscillate about a vertical axis, they turn into the same approximately N–S direction at a given point on the earth's surface: one end is a N-seeking or N pole and the other a S pole of the magnet. The second effect is their attraction for pieces of iron and certain other metals, and in this they are distinguished from the similar attraction exerted by charged bodies (section 1.1) which is independent of the material of which the bodies are composed.

The poles of two magnets exert forces on each other summarized by the qualitative law 'like poles repel, unlike poles attract' (Gilbert, 1540–1603) and they behave in some respects like electric charges but with two important differences. One is that the poles can never be separated, N and S poles always being found in pairs; the second is that conduction of poles never occurs.

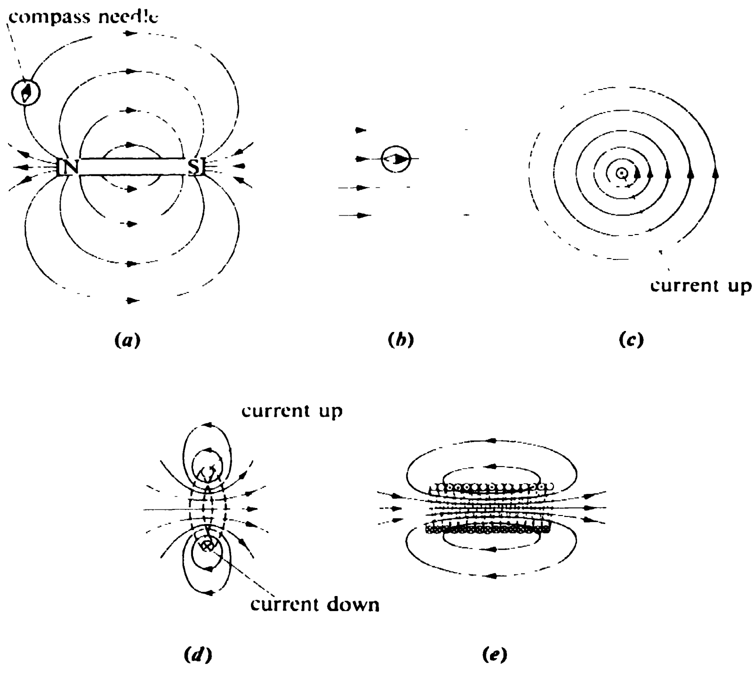
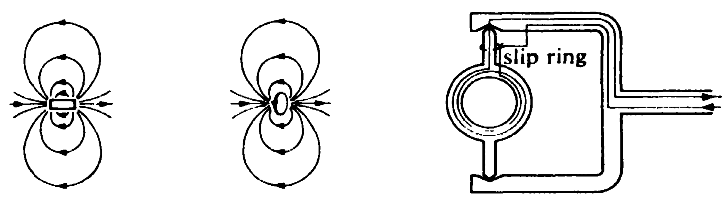


Fig. 7.1. Magnetic lines of force (a) outside a bar magnet, (b) over a small region of the earth's field, (c) outside a long current-carrying wire, (d) due to a circular current-carrying coil, (e) due to a current-carrying solenoid.

Permanent magnets are therefore always *magnetic dipoles* behaving at least qualitatively like electric dipoles in exerting couples on each other. These couples reveal the presence of a *magnetic field* whose direction at a point is obtained by placing a small permanent magnet there and allowing it to come to rest in stable equilibrium: the direction of the S-pole-to-N-pole axis then gives the required direction. This is the 'compass-needle' method of plotting magnetic fields with which readers will be familiar. Magnetic lines of force for a bar magnet and for the earth's field over a small region are shown in Figs. 7.1a and b, the arrows giving the direction a N pole would point.

Current and Magnet. In 1819 Oersted discovered that a current-carrying wire exerted a couple on a nearby compass needle and this, by our method above, means that this is another source of magnetic field. Figures 7.1c, d and e show three examples of fields due to currents. A magnet brought near to a current also exerts a force or couple on it, as we expect from Newton's laws of motion, so the action is a mutual one. Further, the similarity between the lines of force outside the bar magnet and the solenoid suggest that a small solenoid would do as well as a small magnet for detecting and plotting magnetic fields, and this is borne out by simple experiments.

When a permanent dipole is very small its field is indistinguishable from that of a small current loop regarded as a contraction of either Figs. 7.1d or e.' Figure 7.2 shows the similarity and also



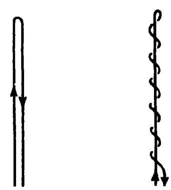
(*a*) (*b*) (*c*)

***Fig. 7.2.** (a) Magnetic field of small permanent magnet; (b) magnetic field of small current loop; (c) a pivoted current-carrying coil which could be used for field-plotting instead of a compass needle.*

shows the form a small pivoted coil can take if we wish to use one for exploring magnetic fields. *The positive direction of the axis of the current loop is related to the direction of current flow by the right-hand screw rule* (appendix 7.1).

Current and Current. It cannot be argued that because currents exert forces on magnets they will exert forces on each other for, as Ampère pointed out, pieces of soft iron attract magnets but do not affect each other. The experiment described in section 1.2 does, however, demonstrate that current–current forces exist and because this type of force would be responsible for the action of the solenoid of Fig. 7.1e on the small coil of Fig. 7.2c used for plotting, we are justified in describing this also as a magnetic force, as indeed we did in chapter 1.

Effect of Leads. In any experiment involving current-carrying circuits the leads are a complication which cannot be avoided. It turns out, however, that a wire doubled back on itself and carrying a steady current produces no magnetic effect even if the return wire has small sinuosities in it (Fig. 7.3). This was originally established



(*a*) (*b*)

***Fig. 7.3.** (a) Wire doubled back on itself; (b) the same with sinuosities. Neither have external magnetic effects if the wires are close.*

by Ampère himself and means that leads play no part in magnetic effects provided they follow closely the same paths to and from the circuits.

7.2 Magnetic Dipoles

We have established so far the following qualitative results: (1) that small permanent magnets and small current loops produce indistinguishable magnetic fields at large distances, (2) that they both behave like ideal dipoles by analogy with the electric dipole and (3) that either may be used to establish the direction of a magnetic field. None of these justifies the assumption of any quantitative laws by analogy with electrostatics: for that we need further experiments.

Figure 7.4 shows a pivoted dipole at A of either the permanent magnet or current-loop type with two other dipoles of either type at

B and C. The distances r_1 and r_2 are large compared with the size of any dipole. In the absence of B and C, A comes to rest with its axis in the direction of the earth's field and we then arrange B and C at distances such that the respective couples exerted on A cancel so that its axis remains in the earth's field. For a given pair of dipoles it is found that the ratios r_1^3/r_2^3 are constant whether the end-on fields are acting, as in Fig. 7.4a, or the broadside, as in Fig. 7.4b. The electric analogy is now quantitative and we can compare

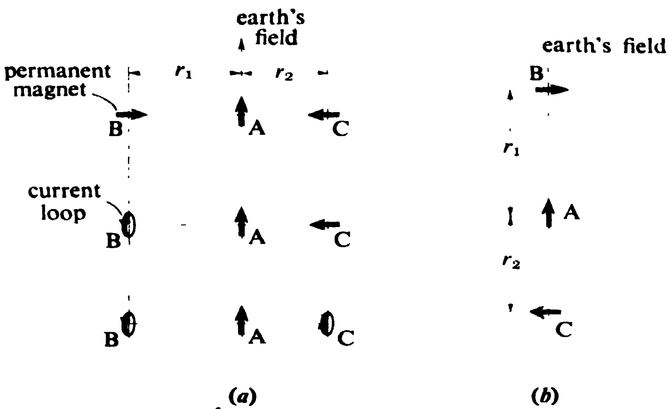


Fig. 7.4. (a) Comparison of dipole moments in end-on positions; (b) comparison in broadside position.

magnetic dipole moments m by comparing the cubes of the distances from A because at the null point $m_1/r_1^3 = m_2/r_2^3$. Once we have selected a standard m the method enables us to measure any other.

If the dipole B in Fig. 7.4a is turned so that its axis makes an angle θ with r_1, it is found that the field at A is such that the effective moment of B is only $m_1 \cos\theta$. This result is consistent with the assumption that magnetic moment may be resolved like a vector quantity, for then the $m_1 \cos\theta$ component would produce a couple on A while the $m_1 \sin\theta$ component would not. A similar result is obtained with the arrangement of Fig. 7.4b. A magnetic dipole moment thus behaves as a vector quantity and is denoted as such by $\mathbf{m}$.

Magnetic Dipole Moment of a Current Loop. By using various current-loop dipoles we find that the magnetic dipole moment is proportional both to the area A and to the current I, the latter

because we have agreed to use the magnetic effect to measure I. Experiment thus shows that m for a current loop is proportional to IA and lies along the positive axis of the loop, i.e. $\mathbf{m} = kI\mathbf{A}$ where k depends on the unit chosen for m.

We choose unit moment as that of a small one-turn loop of unit area carrying unit current. This means that we can summarize all the experiments with current loops by

$$\mathbf{m} = I\mathbf{A} \tag{7.1}$$

as in Fig. 7.5, **m** and **A** both being axial vectors (appendix 7.1), the positive direction of **A** depending on the sense of the current flowing. The equation (7.1) carries the implication that current loops behave like the dipoles of chapter 4 insofar as the dependence of their interaction on distance and angle are concerned.

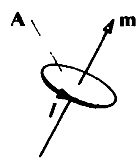


Fig. 7.5. A current-loop dipole.

The MKSA unit for **m** will be the A-m^2. Appendix 7.3 discusses alternative choices for the constant k above, but readers are advised to leave this until a second reading.

7.3 Magnetic Induction or Flux Density

The experiments in sections 7.1 and 7.2 mean that the laws on which we base a study of magnetism are in fact contained in the dipole formulae of Fig. 4.22 coupled with equation (7.1). The laws are so complex that we only use them directly in the case of two small dipoles: to apply them to large magnets and large circuits requires the introduction of field concepts at once.

Definition of **B** *at a Point.* We have already seen that either type of dipole may be used to plot magnetic fields, and now that we have a test dipole of known moment we can formally define a quantity known as *magnetic induction* or *magnetic flux density* (preferring the latter term) denoted by **B**. The *direction* of **B** at a point is the direction of the magnetic moment of a free current-loop dipole in stable equilibrium at the point. The *magnitude* of **B** is the couple

exerted per unit moment on a dipole at right angles to the direction of **B**, or

$$T = m_{\perp}B \qquad \text{(Definition of } B) \quad (7.2)$$

If the angle between **m** and **B** is α (Fig. 7.6) instead of 90°, the vector nature of **m** means that only the $m \sin \alpha$ component experiences a couple and it follows from (7.2) that this couple is

$$\left.\begin{aligned} T_{\alpha} &= mB \sin \alpha \\ \text{or} \qquad \mathbf{T} &= \mathbf{m} \times \mathbf{B} \end{aligned}\right\} \qquad (7.3)$$

using the vector product (appendix 7.2), the latter form including the directions. Equation (7.3) follows directly from the definition of **B** and provides in principle a method of measuring it: in practice, other methods described in chapters 8 and 9 are used.

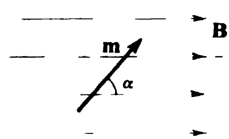


Fig. 7.6. Magnetic dipole at an angle α to **B**.

Equation (7.3) shows that the MKSA unit for **B** is the N/A-m. This is not used, any more than the N/C is for **E** (section 3.1), and the reader should be able to show that an equivalent unit is the V-s/m². The volt-second is known as the *weber*, symbol Wb, so that the unit for **B** usually quoted is the Wb/m² (see section 7.4 for the unit of flux and section 7.5 for a note on CGS units).

The combination of (7.1) and (7.3) gives

$$T_{\alpha} = IAB \sin \alpha \quad \text{to decrease } \alpha \qquad (7.4)$$

for the couple on a small current loop in a magnetic flux density B. Note the similarity between (7.3) and (4.25).

B *at a Point due to a Current Loop.* The definition of **B** coupled with the experiments of section 7.2 show that the flux density due to a current-loop dipole is, using the notation of Fig. 7.7,

$$B_r = \frac{k\,2IA \cos \theta}{r^3}; \qquad B_{\theta} = \frac{k\,IA \sin \theta}{r^3}; \qquad B_{\phi} = 0 \qquad (7.5)$$

where k is a constant determined by the units used. This constant plays the same part as that in Coulomb's law and, for reasons

similar to those given in section 2.4 concerning ε_0, this k is written as $\mu_0/4\pi$, where μ_0 is known as the *permeability of free space*, the *magnetic constant* or just as '*mu nought*'. Equations (7.5) thus become

$$B_r = \frac{\mu_0}{4\pi}\frac{2IA\cos\theta}{r^3}; \qquad B_\theta = \frac{\mu_0}{4\pi}\frac{IA\sin\theta}{r^3}; \qquad B_\phi = 0 \quad (7.6)$$

In the MKSA system, units are already chosen and μ_0 is thus fixed: we shall see in section 8.6 that the choice of the ampere as the unit of current, defined as in section 1.3, makes $\mu_0 = 4\pi \times 10^{-7}$.

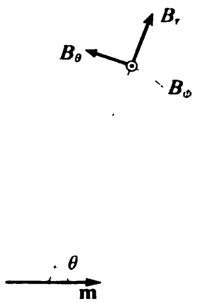


Fig. 7.7. Magnetic flux density due to a magnetic dipole.

Superposition. That two magnetic fields in different directions at a point superpose is easily checked by experiment and justifies the combination of B_r and B_θ in (7.5) into a resultant.

Permanent Magnet Dipoles. The above formulae were expressly related to current-loop dipoles and we have yet to formulate the laws for small permanent magnet dipoles. We agree to use the same unit for magnetic moment as for a current-loop dipole and in that case (7.3) will still apply:

$$\mathbf{T} = \mathbf{m} \times \mathbf{B} \tag{7.7}$$

and the magnetic moment is thus also the couple exerted on a dipole when its magnetic axis is perpendicular to a unit magnetic flux density, a familiar definition. (Note particularly, however, that the magnetic axis of any system can only be found by allowing it to come to rest in a **B** whose direction is already known from current-loop experiments.)

We thus see that small dipoles, whether current loops or permanent magnets, are indistinguishable at points outside them, and

it follows that the magnetic flux density of a permanent magnet dipole of moment m is, at large distances, also

$$B_r = \frac{\mu_0}{4\pi}\frac{2m\cos\theta}{r^3}; \qquad B_\theta = \frac{\mu_0}{4\pi}\frac{m\sin\theta}{r^3}; \qquad B_\phi = 0 \qquad (7.8)$$

in the notation of Fig. 7.7.

7.4 Magnetic Flux

In appendix 4.1 we saw that the flux of any vector **A** over a surface S is $\iint_S \mathbf{A}\cdot d\mathbf{S}$. Normally, in writing, the vector whose flux is in question should be specified, e.g. flux of **E** (in Gauss's theorem), flux of **J** (=current), etc., but the flux of **B** is usually referred to simply as *magnetic flux* and given a special symbol Φ. Thus

$$\Phi_S = \iint_S \mathbf{B}\cdot d\mathbf{S} \qquad \text{(Definition of } \Phi\text{)} \qquad (7.9)$$

defines magnetic flux, the MKSA unit, from the previous section, being the *weber*, Wb.

It will be important in chapter 9 to allocate a sign to magnetic flux crossing the area bounded by a current-carrying circuit: it is counted as positive if it threads the circuit in a direction related to the current by the right-hand screw rule (cf. the rule for **m**, Fig. 7.5): thus the flux is positive when it is in the same direction as the self-flux.

If we use the idea introduced with **E** that lines of force should be drawn so that the number crossing unit area is proportional to the magnitude of **B**, we see that Φ_S will represent the total number of lines crossing the area S.

7.5 Monopoles

Many readers will be used to the idea of magnetic monopoles, or simply poles, as an extension of an analogy between permanent magnets and electric dipoles. The idea clearly arises from the form of the lines of force in Fig. 7.1a which seem to locate separate N and S poles near the ends of the magnet. This concept has its uses in calculations and because of its familiarity this section is devoted to it.

Pole Strength. Accepting that a permanent dipole may be replaced by two poles separated by a displacement **l**, we define pole

strength P so that

$$\mathbf{m} = P\mathbf{l} \qquad \text{(Definition of } P) \quad (7.10)$$

gives the moment of the dipole. The MKSA unit of P will be the A-m.

We know that N or positive poles tend to move in the direction of **B** and S or negative poles in the opposite direction, so that when a dipole is placed in a magnetic field as in Fig. 7.8 and experiences a

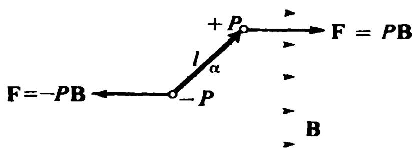


Fig. 7.8. Magnetic dipole as two equal and opposite monopoles.

couple $mB \sin \alpha$ (from (7.7)) or $PlB \sin \alpha$, the force on each pole must be

$$\mathbf{F} = P\mathbf{B} \qquad (7.11)$$

and **B** is thus the force per unit pole

The field due to a dipole is given by (7.8) and, if it originates from a central field due to two unlike equal poles, the field due to a single pole must be

$$B_{\text{pt pole}} = \mu_0 P/4\pi r^2 \qquad (7.12)$$

Equations (7.11) and (7.12) combine to give a law of force between poles

$$F = \mu_0 P_1 P_2/4\pi r^2 \qquad (7.13)$$

in magnitude, a law which can be approximately verified directly using long ball-ended magnets so that the poles are more accurately located.

The CGS e.m.u. System of Units. An important unit system has in the past been based on (7.13) by using CGS mechanical units and defining the unit of pole strength so that $\mu_0/4\pi = 1$ (cf. the e.s.u. system in section 2.4). This is the CGS electromagnetic unit (e.m.u.) system. In it, the unit of current is defined so that it produces a specified magnetic field and the units of **B** and Φ respectively are the *gauss* and the *maxwell*: the unit of current is also

sometimes called the *abampere* or the *biot* (symbol Bi). All our formulae may be converted to e.m.u. by the substitution $\mu_0 = 4\pi$.

Readers familiar with this system will probably have learnt magnetism starting from (7.13): they should have no difficulty in retracing the steps and arriving at all the previous results of this chapter if they so wish. For more discussion of units in general see section 16.7, but appendix 7.3 discusses several alternative forms of the results of this section.

Strength and Limitations of the Pole Concept. The advantage of using monopoles is in the mathematical analogy with electric charge which enables us to take over without further argument some results from chapters 2, 3 and 4. There is, however, no analogy to the conductor of charge so that the results which can be used are somewhat limited. Further, only dipoles exist (i.e. the law of conservation of poles states simply that $\Sigma\, P = 0$ always) and although for the purpose of calculation we may talk of a distribution of, say, north poles, we must always remember that actual systems consist of dipoles (a recent search for magnetic monopoles predicted by Dirac has had no success—see CERN, 1923). It will become clear in chapter 14 that a magnet is more akin to a polarized dielectric than to a charged conductor.

The limitations of poles are first that practical *permanent* dipoles do not exist: the moment of a magnet is affected by any nearby magnets and even by the earth's field (although isolated elementary particles such as electrons or neutrons are permanent dipoles). One reason why we prefer to use current loops for the experimental basis is that their moment can be maintained constant. Secondly, the location of the pole and thus, by (7.10), its strength, are indefinite and only the magnetic moment is an accurately measurable quantity. Thirdly, evidence which we shall consider in chapter 14 all points to the association of angular momentum with magnetic moment and the current loop is more appropriate. Nevertheless, the pole can be used to the full in developing electromagnetic theory (Katz, 1962).

7.6 Relations involving Permanent Magnet Dipoles

As long as the limitations just enumerated are remembered, many of the results from chapters 3 and 4 may be taken over for use with small permanent magnets (and with small current loops, except for (7.14) below, which was derived on the basis of two separate charges or poles: it is justified for current loops in section 8.8).

Further to the dipole formulae above, we have the force on a dipole in a non-uniform magnetic field (cf. equation 4.27):

$$F_x = m_x\,\partial B_x/\partial x + m_y\,\partial B_x/\partial y + m_z\,\partial B_x/\partial z, \text{ etc.} \tag{7.14}$$

and the potential energy of an already-established dipole in a magnetic flux density (cf. (4.28))

$$U = -\mathbf{m}\cdot\mathbf{B} \tag{7.15}$$

while the dipole field may be expressed either as in (7.8) or as

$$B_x = \mu_0 m(2x^2 - y^2)/4\pi r^5$$

$$B_y = \mu_0 3mxy/4\pi r^5 \quad \text{where} \quad r = (x^2+y^2)^{1/2} \tag{7.16}$$

from problem (4.12), or as

$$\mathbf{B} = \frac{3\mu_0 m\cos\theta}{4\pi r^4}\,\mathbf{r} - \frac{\mu_0}{4\pi r^3}\,\mathbf{m} \tag{7.17}$$

from problem 4.14.

Magnetometers. For comparing magnetic moments or fields various forms of magnetometer are available. The *deflection magnetometer* is the instrument of Fig. 7.4: if the horizontal component of the earth's magnetic flux density is B_0 then the magnet A, if acted on by only one other field at right angles to B_0, turns through an angle θ given by

$$B = B_0 \tan\theta \tag{7.18}$$

equating the opposing couples.

Alternatively, a magnet suspended from its centre point in a horizontal plane in which a magnetic flux density B occurs experiences a couple $mB\theta$ if given a *small* angular displacement θ from the position of stable equilibrium. If the moment of inertia of the magnet about an axis through the suspension is I_m then the equation of motion is

$$I_m\,d^2\theta/dt^2 = -mB\theta \tag{7.19}$$

from which the angular motion is harmonic with a period

$$T_0 = 2\pi\sqrt{\frac{I_m}{mB}} \tag{7.20}$$

for small oscillations, assuming a torsionless suspension. This is the *vibration magnetometer*.

A third possibility is to arrange a magnetic field to act at right angles to a magnet suspended by a fibre or between two vertical fibres of torsion constant c. The deflection produced is annulled by a twist of the torsion head through θ producing a couple $c\theta$ which balances mB due to the magnetic field. Thus

$$B = c\theta/m \tag{7.21}$$

This is the *torsion magnetometer.*

In all these, B is a function of constants of the instruments and a variable (θ or T_0) which is thus a measure of B.

Magnetic Potential and Magnetomotive Force. By analogy with electrostatics it is possible to introduce a magnetic potential difference defined by the relation $V_B - V_A = \int_A^B -\mathbf{B}\cdot d\mathbf{s}$. For historical reasons, however, it is conventional to define

$$V_m = V_B - V_A = \int_A^B -\frac{\mathbf{B}\cdot d\mathbf{s}^*}{\mu_0} \quad \text{(Definition of } V_m\text{)} \tag{7.22}$$

as the magnetic scalar potential (since in more advanced treatments there is also a *vector* potential). It follows as in the corresponding electrostatic case in section 3.6 that the component of $\mathbf{B}$ in any direction $\mathbf{s}$ is given by

$$B_s = -\mu_0 \frac{\partial V_m}{\partial s} \tag{7.23}$$

with expressions similar to (3.20) and (3.21) for particular coordinate systems. We shall have, for instance, for a dipole of moment m

$$V_m = \frac{m \cos \theta}{4\pi r^2} \qquad \text{(zero at } \infty\text{)} \tag{7.24}$$

from which $B_r = (\mu_0 2m \cos \theta)/4\pi r^3$, etc. as in equation (7.8).

* At this stage it is sometimes customary to introduce the quantity $\mathbf{H} = \mathbf{B}/\mu_0$, the *magnetic field strength in vacuo*, so that magnetic potential difference is $\int \mathbf{H}\cdot d\mathbf{s}$, etc. Because we shall have to redefine it inside magnetic media in chapter 14, there seems little point in introducing it here. Readers who wish to work in terms of $\mathbf{H}$ can easily do so, but for conversion to CGS e.m.u. they should consult appendix 14.1 The MKSA unit of $\mathbf{H}$ is the A/m; the CGS e.m.u. is the *oersted.* Note also that to convert formulae containing V_m to e.m.u., the substitution $V_m \rightarrow V_m/4\pi$ must also be made (section 16.7).

The *magnetomotive force* or m.m.f. $\mathscr{H}$ is defined round a closed path C wholly *in vacuo* by

$$C \text{ in vacuo} \qquad \mathscr{H} = \oint_C \frac{\mathbf{B}\cdot d\mathbf{s}}{\mu_0} \quad \text{(Definition of } \mathscr{H}\text{)} \qquad (7.25)$$

by analogy with e.m.f. Clearly for a distribution of poles, $\mathscr{H}$ round a closed path is zero, just as $\mathscr{E}=0$ in electrostatics. Thus

$$C \text{ in vacuo; poles only} \qquad \oint_C \frac{\mathbf{B}\cdot d\mathbf{s}}{\mu_0} = 0 \qquad (7.26)$$

Because $\mathbf{B}$ is the force per unit pole, potential differences and m.m.f.s can be interpreted in terms of the work done in taking unit pole along the paths (divided by μ_0), but this is not essential.

Gauss's Theorem. For a set of permanent magnets, Gauss's theorem will take the form that the magnetic flux over any closed surface wholly *in vacuo* is zero, or

$$S \text{ in vacuo; poles only} \qquad \oiint_S \mathbf{B}\cdot d\mathbf{S} = 0 \qquad (7.27)$$

which expresses the absence of unbalanced poles. Naturally, in hypothetical problems involving one type of pole only, the theorem takes the same form as in electrostatics and is proved in the same way, but (7.27) is applicable to real situations.

7.7 Summary of Chapter 7

We have been able to progress quickly in the study of magnetic forces because of the experience gained with electric ones, though we must be wary of taking the analogy too far. The law of interaction between magnetic dipoles has been split into two parts:

Definition of $\mathbf{B}$: $\quad \mathbf{T} = \mathbf{m}\times\mathbf{B}$.
Value of $B_{r\,\text{dipole}} = \mu_0 2m\cos\theta/4\pi r^3$, etc.

which correspond to the splitting of Coulomb's law (section 3.10):

Definition of $\mathbf{E}$: $\quad \mathbf{F} = Q\mathbf{E}$. $\quad$ Value of $E_{\text{pt. charge}} = Q/4\pi\varepsilon_0 r^2$.

The experiments supporting the magnetic law rest on much cruder foundations than does Coulomb's law: we shall see, however, that some direct conseqences of the laws can be very accurately verified (section 16.2).

Experiment also showed that the magnetic dipole moment of a small current loop can be expressed as

$$\mathbf{m} = I\mathbf{A} \tag{7.1}$$

and we have also defined the magnetic moment of a small permanent magnet such that $\mathbf{T} = \mathbf{m} \times \mathbf{B}$ applies in both cases. Perhaps the most important result is that outside the dipoles we cannot distinguish one type from the other.

The monopole concept has been critically examined and we have seen how many electrostatic relations can be directly taken over (section 7.6).

The *magnetism of permanent magnet dipoles* is summarized by

$$\oiint_S \mathbf{B} \cdot d\mathbf{S} = 0 \tag{7.27}$$

for any surface S lying wholly *in vacuo*, indicating the absence of free poles; together with

$$\oint_C \mathbf{B} \cdot d\mathbf{s} = 0 \tag{7.26}$$

for any path C *wholly in vacuo* (this restriction is necessary since $\mathbf{B}$ has not been defined within the permanent magnets themselves).

Appendix 7.1 Axial and Polar Vectors

Many vector quantities are associated not with a direction in space but with rotation about an axis and if we wish to add or resolve them we find that the axis itself can be used as the direction of the vector (Fig. 7.9). These are known as *axial vectors* as opposed to the *polar vectors* like force and velocity hitherto en-

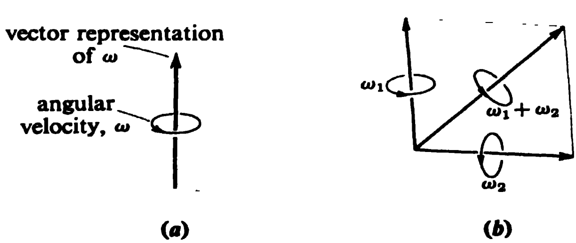


Fig. 7.9. (a) Representation of an axial vector, angular velocity; (b) addition of angular velocities.

countered. Angular velocity, angular momentum and couple or torque are examples of axial vectors.

A convention is adopted that the positive direction of the vector representation is given by the following rule: rotate a right-handed screw in the same sense as that of the quantity and it will proceed in the positive direction along the axis (Fig. 7.9a). An area associated with a rotation as in the current-loop dipole is also an axial vector whose direction is given by the same convention; other areas are also axial vectors whose directions are assigned as in appendix 4.1.

Appendix 7.2 The Vector Product of Two Vectors

Two vectors **A** and **B** as in Fig. 7.10 define a plane, that of the parallelogram between them. The *vector product* of **A** and **B**, denoted by $\mathbf{A}\times\mathbf{B}$ or $\mathbf{A}\wedge\mathbf{B}$ (read 'A cross B'), is defined as a vector whose magnitude is $|A|\,|B|\sin\theta$ and whose direction is normal to the plane containing **A** and **B** and with a positive sense such that **A**, **B** and $\mathbf{A}\times\mathbf{B}$ form a *right-handed system*.

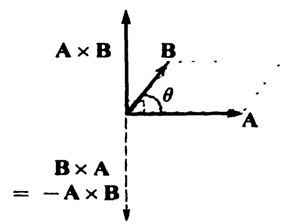


Fig. 7.10. The vector product.

The term 'right-handed system' means that the rotation of a right-handed screw from the first-named quantity (**A**) to the second (**B**) through the smaller angle between the positive directions of **A** and **B** (θ) proceeds in the direction of the third quantity ($\mathbf{A}\times\mathbf{B}$) as in Fig. 7.10. (Most texts use a right-handed system of cartesian co-ordinates x, y, z.)

While this appears complex at first sight, it is well worth grasping for the simplification of notation it brings later. We have used it in this chapter in equation (7.3) and it should be checked that $\mathbf{m}\times\mathbf{B}$ does give an axial vector **T** in the correct direction.

We quickly have that $\mathbf{A}\times\mathbf{B}=-(\mathbf{B}\times\mathbf{A})$ (commutation does not apply), that $\mathbf{A}\times\mathbf{B}=0$ if **A** and **B** are parallel, and $\mathbf{A}\times(\mathbf{B}+\mathbf{C})=(\mathbf{A}\times\mathbf{B})+(\mathbf{A}\times\mathbf{C})$. Also note that $AB\sin\theta$ is the area of the parallelogram contained by **A** and **B** so that $\mathbf{A}\times\mathbf{B}$ could be said to represent the area as a vector quantity.

Appendix 7.3 Variations in MKSA Units

Different authors using MKSA units adopt, at the stage reached in this chapter, one of two different definitions of the moment of a current loop—a situation which easily causes confusion. The two choices are designated by the names Sommerfeld and Kennelly as suggested by the Coulomb's Law Committee of the American Association of Physics Teachers (*American Journal of Physics*, 1950), and that adopted in this book is the Sommerfeld. The choice will only affect future work in chapters 11 and 14 and the first part of the appendix deals with this. The second part deals with a further source of confusion which only arises when the monopole is used and is thus confined to section 7.5.

Sommerfeld and Kennelly MKSA Units. All authors agree on the fundamentals, as they must, and all agree to define **B** as couple per unit current loop ($I = 1$, $A = 1$) so that the following are not in dispute:

$$\mathbf{T} = I\mathbf{A}\times\mathbf{B} \quad \text{and} \quad B_r = \frac{\mu_0}{4\pi}\frac{2IA\cos\theta}{r^3}$$

The rest is best displayed in two columns remembering (footnote to section 7.6) that by **H** is meant $\mathbf{B}/\mu_0$.

Table 7.1

	Sommerfeld	*Kennelly*
Moment of current loop	$\mathbf{m} = I\mathbf{A}$	$\mathbf{m} = \mu_0 I\mathbf{A}$
Hence	$\mathbf{T} = \mathbf{m}\times\mathbf{B}$	$\mathbf{T} = \mathbf{m}\times\mathbf{B}/\mu_0 = \mathbf{m}\times\mathbf{H}$
and	$B_r = \dfrac{\mu_0}{4\pi}\dfrac{2m\cos\theta}{r^3}$, etc.	$B_r = \dfrac{2m\cos\theta}{4\pi r^3}$, etc. and $H_r = 2m\cos\theta/4\pi\mu_0 r^3$, etc.
Unit of **m**	A-m^2	Wb-m

Thus we have the characteristic feature of Sommerfeld that the μ_0 is in the numerator of **B** and of Kennelly that it is in the denominator of **H**: it is purely a matter of definition, which does not affect formulae relating to currents *in vacuo* (next chapter) but only causes differences when magnetic media are present (chapter 14, appendix 14.1).

Monopole Formulae. Both Sommerfeld (S) and Kennelly (K) define pole strength so that $\mathbf{m} = P\mathbf{l}$ and this leads as we see below to an inverse square law between poles which has μ_0 in the numerator for S and in the denominator for K: not surprising when the different units in which P is expressed are examined. Recently, however, Shire (1960) has shown how the S definition $\mathbf{m} = I\mathbf{A}$ need not lead to μ_0 in the numerator of the pole formula and it seems worthwhile to clear up this matter by indicating all possible choices in four columns as in table 7.2.

Table 7.2

	Sommerfeld	*Shire*	*Kennelly*	*Choice 4*
Moment of current loop	As Sommerfeld $\mathbf{m} = I\mathbf{A}$		As Kennelly $\mathbf{m} = \mu_0 I\mathbf{A}$	
Moment of permanent dipole	$\mathbf{m} = P\mathbf{l}$	$\mathbf{m} = \dfrac{P\mathbf{l}}{\mu_0}$	$\mathbf{m} = P\mathbf{l}$	$\mathbf{m} = \mu_0 P\mathbf{l}$
Hence force on pole P (as (7.11))	$\dfrac{\mathbf{F}}{P} = \mathbf{B}$	$\dfrac{\mathbf{F}}{P} = \dfrac{\mathbf{B}}{\mu_0} = \mathbf{H}$	$\dfrac{\mathbf{F}}{P} = \dfrac{\mathbf{B}}{\mu_0} = \mathbf{H}$	$\dfrac{\mathbf{F}}{P} = \mathbf{B}$
and hence B due to single pole P (as (7.12))	$\mathbf{B} = \dfrac{\mu_0 P}{4\pi r^2}\hat{\mathbf{r}}$	$\mathbf{B} = \dfrac{P}{4\pi r^2}\hat{\mathbf{r}}$	$\mathbf{B} = \dfrac{P}{4\pi r^2}\hat{\mathbf{r}}$	$\mathbf{B} = \dfrac{\mu_0 P}{4\pi r^2}\hat{\mathbf{r}}$
and hence inverse square law between poles (as (7.13))	$\mathbf{F} = \dfrac{\mu_0 P_1 P_2}{4\pi r^2}\hat{\mathbf{r}}$	$\mathbf{F} = \dfrac{P_1 P_2}{4\pi\mu_0 r^2}\hat{\mathbf{r}}$	$\mathbf{F} = \dfrac{P_1 P_2}{4\pi\mu_0 r^2}\hat{\mathbf{r}}$	$\mathbf{F} = \dfrac{\mu_0 P_1 P_2}{4\pi r^2}\hat{\mathbf{r}}$
Unit of P	A-m	Wb	Wb	A-m

Again a matter of definition. It is clearly only in the column headed Sommerfeld that the substitution $\mu_0 = 4\pi$ consistently converts formulae to CGS e.m.u. and this is one of the reasons for adopting it in this book.

Reference

For a recent appraisal of the work of Ampère see Tricker (1962).

PROBLEMS

SECTION 7.1

7.1 A circular steel disc is magnetized uniformly in a direction parallel to one of its diameters. How would you identify this diameter using no other electric or magnetic apparatus?

SECTION 7.2

7.2 Find the magnetic moment of a circular coil of radius 0·5 cm carrying a current of 1 A in 10 turns. What is the force between two such coils arranged coaxially 10 cm apart?

SECTION 7.3

7.3 Assuming that the earth's magnetic field is the same as that of a magnetic dipole at the centre whose axis lies along that of the earth, show that the angle of dip δ at a point whose latitude is λ is given by $\tan \delta = 2 \tan \lambda$.

SECTION 7.5

7.4 Compare and contrast the properties of electric charges and magnetic poles.

SECTION 7.6

*7.5 Two magnetic dipoles are a large distance r apart and their moments m_1 and m_2 are coplanar and make angles θ_1, θ_2 respectively with the positive direction of r. Show, by using (7.8) and (7.15), that the mutual potential energy is $\mu_0 m_1 m_2(\sin \theta_1 \sin \theta_2 - 2\cos \theta_1 \cos \theta_2)/4\pi r^3$. Hence find the forces and couples between dipoles oriented as in Fig. 4.22.

*7.6 A circular disc of radius a is magnetized perpendicular to its plane with a magnetic moment per unit area m. Find the magnetic flux density at a point on its axis a distance x from the centre. (Use (7.16) and see also the next chapter.)

*7.7 A sphere of radius a is magnetized uniformly parallel to a diameter so that its magnetic moment per unit volume is M. Find the magnetic flux density at a distance r from the centre for $r > a$ and $r < a$. (The least laborious method is to treat the sphere of dipoles as two spheres of poles of opposite sign slightly displaced from each other. The flux densities due to each sphere are given by expressions similar to those derived for a sphere of charge—equation (4.12).)

APPENDIX 7.2

7.8 Show that $\mathbf{A}\times\mathbf{B} = \mathbf{i}(A_yB_z - A_zB_y) + \mathbf{j}(A_zB_x - A_xB_z) + \mathbf{k}(A_xB_y - A_yB_x)$ where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors along the x-, y- and z-axes respectively.

7.9 Show that $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ represents the volume of the parallelepiped whose sides are $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Hence show that $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A})$, etc. and $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = -\mathbf{A}\cdot(\mathbf{C}\times\mathbf{B})$, etc.

CHAPTER 8

FORCES BETWEEN STEADY CURRENTS (*IN VACUO*)

The forces and couples between current-carrying circuits whose dimensions are small compared with their distance apart can be calculated from the formulae developed for small permanent magnet dipoles. The interaction between large circuits is likely to be more complex even than that between dipoles, so the problem is again best split into two halves. Thus the first three sections of this chapter derive different methods for finding the magnetic flux density at a point due to a current, while the following two are concerned with the action of a known flux density on currents placed in it. Only after this do we combine the two into a direct law of force between currents and relate it to the experiments and definitions of chapter 1. We end by seeing how the laws might be extended to moving charges in general.

8.1 Magnetic Fields due to Currents 1: Magnetic Shell Equivalence

The leads to the large circuit shown in Fig. 8.1a are known already to have no magnetic effect, so to calculate the value of **B** at P we need only consider the circuit itself. If we introduce a network of wires as in Fig. 8.1b and arrange each mesh to carry a current I in the same sense as that in Fig. 8.1a then because, again, we know that two equal and opposite currents along the same wire produce no magnetic effect, only those mesh currents along the outer wire belonging to the original circuit are unbalanced and produce a field at P. Hence **B** at a point is the same for the system of mesh currents as for the actual circuit.

If the mesh is fine enough for each to have an area dS whose linear dimensions are small compared with the distance r from P then the net acts as a collection of current-loop dipoles each of moment IdS by (7.1) and is called an *equivalent magnetic shell*: for, by the analysis in chapter 7, the fine net would produce the same field at P as a thin sheet of material magnetized with N poles on one side and S poles

on the other whose magnetic moment per unit area (or *strength*) were I.

From (7.24), the magnetic potential at P due to a typical mesh is

$$dV_m = (I\,dS\cos\theta)/4\pi r^2 \qquad \text{(zero at } \infty) \qquad (8.1)$$

$$= I\,d\Omega/4\pi \qquad (8.2)$$

where $d\Omega$ is the solid angle subtended at P by dS. Hence

$$V_m = I\Omega/4\pi \qquad \text{(zero at } \infty) \qquad (8.3)$$

Ω being the total solid angle subtended at P by the whole circuit or shell. The magnetic flux density at P is then obtained from

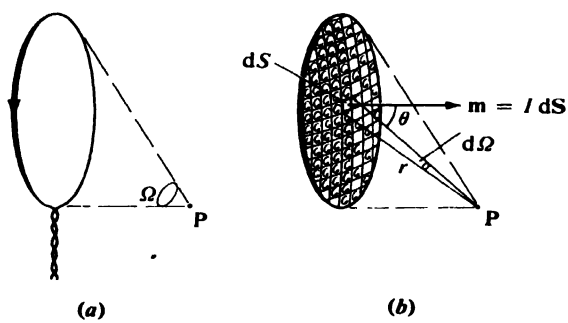


Fig. 8.1. (a) The actual circuit; (b) a net giving the same magnetic effect as the circuit —an equivalent magnetic shell.

$B_x = -\mu_0\,\partial V_m/\partial x$, etc., as in (7.23). To obtain the correct sign for V_m from (8.3)—namely + on the N pole side—we must allocate a sign to Ω by the convention that it is to be reckoned positive at P if the current viewed from P appears to flow in a counterclockwise sense and vice versa.

Notice that the potential depends only on the periphery of the net or shell and on the current or strength, but not at all on the shape, which may therefore bulge as we wish.

Magnetic Flux Density on the Axis of a Circular Coil. As an application of the above method, we find B on the axis of a one-turn circular coil of radius a carrying a current I. The solid angle Ω subtended at P (Fig. 8.2) a distance x along the axis from the

centre is $2\pi(1-\cos\theta)$ by appendix 2.3. Using (8.3)

$$V_m = \frac{I}{2}\left(1 - \frac{x}{(x^2+a^2)^{1/2}}\right)$$

and thus

$$B_x = -\mu_0\,\partial V_m/\partial x = \frac{\mu_0 I a^2}{2(x^2+a^2)^{3/2}}; \qquad B_y = 0; \qquad B_z = 0 \quad (8.4)$$

B_x is plotted against x in Fig. 8.2b. When $x=0$, at the centre,

$$B_x = \mu_0 I/2a \quad (8.5)$$

while as $x\to\infty$, $B_x\to\mu_0 Ia^2/2x^3$ or $\mu_0 2IA/4\pi x^3$ where A is the area: this is consistent with (7.6).

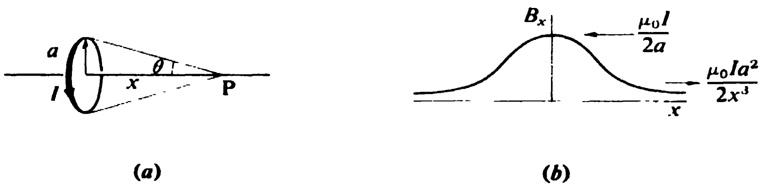


Fig. 8.2. The magnetic field of a circular coil.

A simpler method of obtaining (8.4) will be developed in section 8.3: meanwhile we look at two important applications of the expression.

Helmholtz Coils. If two circular coils of the same radius and carrying identical currents in the same sense are arranged coaxially their flux densities along the axis will add. A highly uniform field

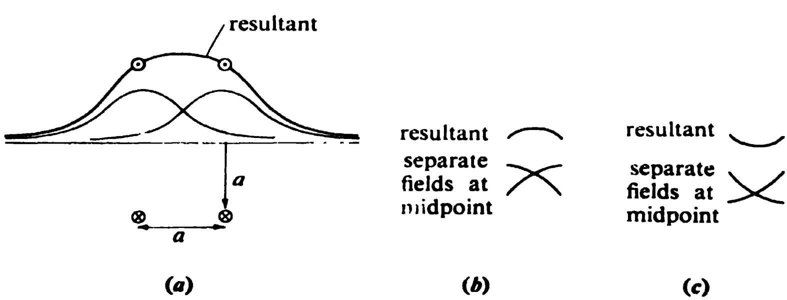


Fig. 8.3. (a) Helmholtz coils; (b) coils closer than (a); (c) coils further apart than (a).

will be produced at the axial point midway between the coils if the decrease in one field as we move away from this point is exactly compensated by the increase in the other. This condition occurs when the variation of B_x with x at the midpoint is linear, i.e. where $d^2B_x/dx^2=0$. Applied to equation (8.4), this gives $x=\frac{1}{2}a$ and the coils should therefore be placed a distance apart equal to the radius of either, an arrangement known as *Helmholtz coils* (Fig. 8.3).

B on the Axis of a Solenoid. If a solenoid of n turns per unit length is closely wound it may be regarded as a collection of n coaxial coils per unit length and an element of length dx as a coil of $n\,dx$ turns. At a point P on the axis, let the angles subtended by the ends of the solenoid be α and β as in Fig. 8.4 and that subtended

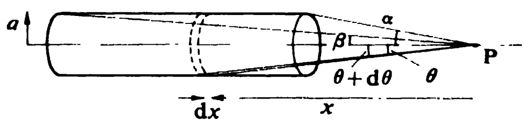


Fig. 8.4. Calculation of the magnetic flux density due to a solenoid.

by the typical element be θ. For a current I, using (8.4),

$$dB_P = \mu_0 n\,dxIa^2/2(x^2+a^2)^{3/2}$$

and because $x=a\cot\theta$ and $dx=-a\operatorname{cosec}^2\theta\,d\theta$,

$$dB_P = -\mu_0 nI\sin\theta\,d\theta/2$$

Integrated from α to β this gives

$$B_P = \frac{\mu_0 nI}{2}(\cos\beta-\cos\alpha) \tag{8.6}$$

For a solenoid whose length becomes much greater than its radius, $\alpha\to\pi$ and $\beta\to 0$ and in the limit

$$B_P = \mu_0 nI \tag{8.7}$$

for an infinite solenoid ($4\pi nI$ in CGS e.m.u. by the substitution $\mu_0=4\pi$). Examination of (8.6) for solenoids with length much greater than a shows that (8.7) is a good approximation to B over a fair region in the centre and gives a further method of obtaining a nearly uniform field.

8.2 Magnetic Fields due to Currents 2: The Circuital Theorem

The magnetic potential difference between two points A and B in the magnetic field of a current is obtained by applying (8.2), i.e. by finding the *change* in solid angle along a path from B to A and multiplying by $I/4\pi$. In Fig. 8.5, Ω_A and Ω_B are the magnitudes of the solid angles and by our sign convention that at B must be $-\Omega_B$. Along path C_1 the solid angle increases from $-\Omega_B$ to zero and then

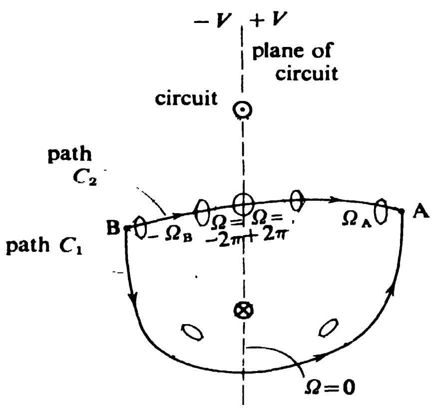


Fig. 8.5. Derivation of the circuital theorem.

increases further to $+\Omega_A$, a total *increase* of $\Omega_A+\Omega_B$. Hence

$$V_A - V_B = I(\Omega_A+\Omega_B)/4\pi \quad \text{along } C_1 \tag{8.8}$$

Along path C_2, the solid angle first decreases from $-\Omega_B$ to -2π and then, because we change sides, decreases further from $+2\pi$ to Ω_A, a total *decrease* of $4\pi-(\Omega_A+\Omega_B)$. Hence

$$V_A - V_B = I(\Omega_A+\Omega_B-4\pi)/4\pi \quad \text{along } C_2 \tag{8.9}$$

By choosing other paths which encircle the current, an infinite set of values for the magnetic potential difference can be found, all differing from (8.8) by a multiple of I: *the magnetic potential due to a current is not path-independent or single-valued*, a fact we might have suspected from (8.3) since Ω cannot be distinguished from $\Omega+4\pi$, etc. Because $V_A - V_B$ is defined as $\int_B^A -\mathbf{B}\cdot d\mathbf{s}/\mu_0$, it follows that $\oint \mathbf{B}\cdot d\mathbf{s}/\mu_0$ or the m.m.f. round a closed path is not necessarily zero.

In fact, if we traverse C_1 in Fig. 8.5 from B to A, and then C_2 from A to B, thus completing a closed path, the contribution to the m.m.f. by (8.8) and (8.9) is just I. Thus

$$\oint_C \mathbf{B}\cdot d\mathbf{s} = \mu_0 I \tag{8.10}$$

for a closed path C encircling a current I once only. This is known as the *circuital theorem* and the path C we shall call an Amperian path. Another way of expressing the theorem is to say that the work done in taking a unit pole round any closed path C is μ_0 multiplied by the current encircled by C.

The use of (8.10) is similar to that of Gauss's theorem in electrostatics: its main value lies in summarizing the properties of magnetic fields and in the development of theory. As a method of calculating **B** it is restricted to highly symmetrical cases where the Amperian path can usually be chosen so that **B** and ds are parallel and B is constant. The left-hand side is then simply the product of B and the path length.

Although not single-valued, magnetic potential difference does not become useless because the 4π's which can be added to it do not enter B after differentiation. Moreover, if we restrict any paths by imagining a barrier across the circuit carrying the current, the m.m.f. always $=0$ and V_m becomes single-valued. This is clearly impossible inside a mass of material carrying a volume distribution of current and here V_m can never be used.

B due to a Long Straight Current-Carrying Wire. A straight wire

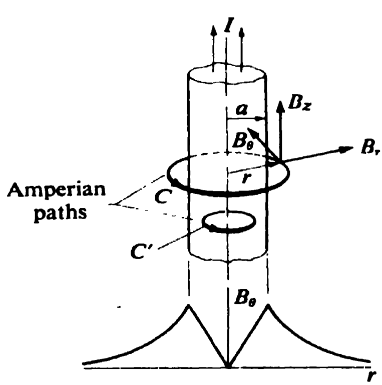


Fig. 8.6. Magnetic flux density due to a long straight current-carrying wire.

of infinite length and circular section of radius a carrying a steady current I (Fig. 8.6a) has not only cylindrical symmetry about the axis but constancy along the axis as well. Thus to obtain B at P a distance r from the axis outside the wire, denote by B_θ the component along the Amperian path C in the figure. B_θ is the same at all points on C by the symmetry, and the circuital theorem therefore gives

$$2\pi r B_\theta = \mu_0 I$$

or, $r \geqq a$
$$B_\theta = \frac{\mu_0 I}{2\pi r} \tag{8.11}$$

If P is inside the wire and the current is distributed uniformly over the cross-section, the path C' encircles a current Ir^2/a^2 and

$$2\pi r B_\theta = \mu_0 I r^2/a^2$$

or, $r \leqq a$
$$B_\theta = \frac{\mu_0 I r}{2\pi a^2} \tag{8.12}$$

The variation of B_θ with r is plotted in Fig. 8.6b. Gauss's theorem (section 8.8) shows that B_r and B_z must be zero.

B due to a Toroidal Coil. Applying the circuital theorem to the Amperian path C in Fig. 8.7 within a closely wound toroidal coil with n turns per unit length gives

$$2\pi r B = \mu_0 2\pi a n I$$

or
$$B = \mu_0 n I \, a/r \tag{8.13}$$

so that on the axis where $r = a$,

$$B = \mu_0 n I \tag{8.14}$$

as for an infinite solenoid.

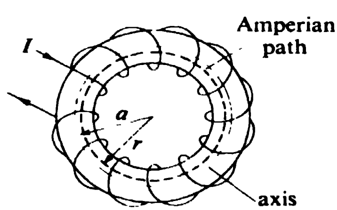


Fig. 8.7. A toroidal coil.

8.3 Magnetic Fields due to Currents 3: The Current Element Formula

Figure 8.8 shows a small length dl of a circuit carrying a current I: this is known as a current element. We seek a formula for the flux density dB at P in terms of I, dl, θ and r which will yield correct results for the total flux density at P when summed round the complete circuit of which dl is a part.

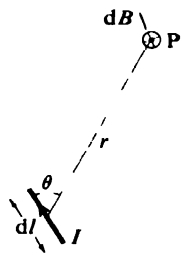


Fig. 8.8. A current element and its relation to the point P.

To find such a formula we choose an equivalent magnetic shell or mesh of currents of such a shape that with each element of the circuit we can associate part of the shell. Such a shape is the 'basket-ball' net of Fig. 8.9 which has been pushed out until

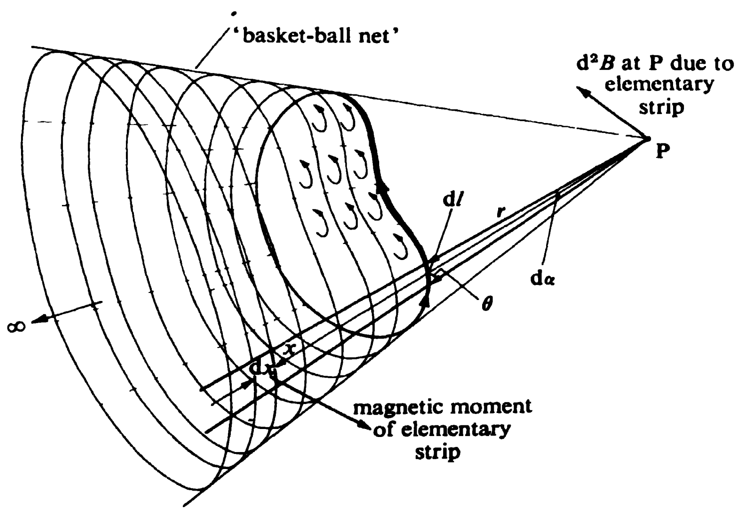


Fig. 8.9. Derivation of the current element formula for the magnetic flux density at P.

straight lines radiating from P are generators of it. The net extends to infinity and associated with any element of the circuit dl is a strip cut off by the generators from P as shown. The strip is divided into elements, a typical one being a distance x from P, of width dx and length $x\,d\alpha$ where $d\alpha$ is the plane angle subtended at P by dl. The element of the strip thus has an area $x\,dx\,d\alpha$ and hence a magnetic moment broadside to P of $Ix\,dx\,d\alpha$. The magnetic flux density at P due to the strip is

$$d^2B = \frac{\mu_0}{4\pi}\,\frac{Ix\,dx\,d\alpha}{x^3}$$

from (7.6), in a direction normal to the plane of dl and r as shown. The whole strip thus produces

$$dB = \int_{x=r}^{x=\infty} \mu_0 I\,dx\,d\alpha/4\pi x^2$$
$$= \mu_0 I\,d\alpha/4\pi r$$

and, since $d\alpha = (dl \sin\theta)/r$,

$$dB = \frac{\mu_0 I\,dl \sin\theta}{4\pi r^2} \tag{8.15}$$

in a direction at right angles to dl and r such that the right-handed screw rule is obeyed. This is the required formula because the shell can be closed by a spherical cap at infinity whose field at P can be shown to be zero (problem 8.7) Equation (8.15) can be put into vector form so that the directions are included:

$$\mathbf{dB} = \frac{\mu_0 I\,\mathbf{dl} \times \hat{\mathbf{r}}}{4\pi r^2} = \frac{\mu_0 I\,\mathbf{dl} \times \mathbf{r}}{4\pi r^3} \tag{8.16}$$

in which an extra r has been introduced in the numerator and denominator and $\hat{\mathbf{r}}$ is a unit vector along $\mathbf{r}$. Figure 8.8 shows the directions.

This formula, sometimes called the Biot Savart law and sometimes Laplace's formula, cannot be verified experimentally in a direct way because an element can never produce a field without the rest of the circuit doing so when currents are steady; indeed the formula strictly has no meaning except when used for a complete circuit and any addition to it which disappeared on integration would make an equally valid current element formula.

B due to a Finite Wire. Using the notation of Fig. 8.10, a typical element produces a flux density at P

$$dB = \frac{\mu_0 I \, dl \sin\theta}{4\pi(r^2+l^2)}$$

at right angles to the plane of the diagram. Because $l = r \cot\theta$ and $dl = -r \operatorname{cosec}^2\theta \, d\theta$,

$$dB = -\mu_0 I \sin\theta \, d\theta / 4\pi r$$

and integration from α to β yields

$$B = \frac{\mu_0 I}{4\pi r}(\cos\beta - \cos\alpha) \tag{8.17}$$

For a very long wire, $\alpha \to 0$ and $\beta \to \pi$ and in the limit $B = \mu_0 I / 2\pi r$ which checks with (8.11). Equation (8.17) can be used to determine

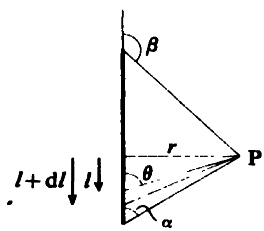


Fig. 8.10. Magnetic flux density due to a finite current-carrying wire.

fields due to circuits which are combinations of finite wires such as rectangular coils. (See problem 8.6.)

A Circular Coil. Although this problem has already been solved in section 8.1 it is instructive to reconsider it. In Fig. 8.11 the flux densities due to two elements at opposite ends of a diameter are

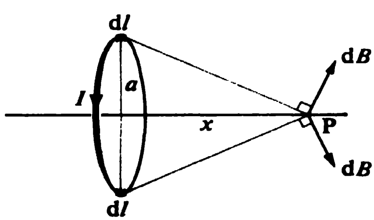


Fig. 8.11. Derivation of flux density due to a circular coil using the current element formula.

shown, each being equal to $\mu_0 I \,dl/4\pi(x^2+a^2)^{1/2}$. The components perpendicular to the axis cancel and those along the axis add and, since this applies to every pair of elements, the total B at P is simply the sum of all the resolved parts along x. The reader will easily verify that this yields (8.4).

8.4 Currents in Magnetic Fields 1: Magnetic Shell Equivalence

If a current doubled back upon itself exerts no magnetic forces it can, by Newton's third law of motion, experience none. It follows that the forces and couples on a current-carrying circuit are the same as those on the equivalent magnetic shell because the internal net introduced carries only equal and opposite currents.

One straightforward application of this is to a plane circuit of area A carrying a current I in a uniform magnetic field of flux density B (Fig. 8.12a). If the normal to the plane makes θ with B

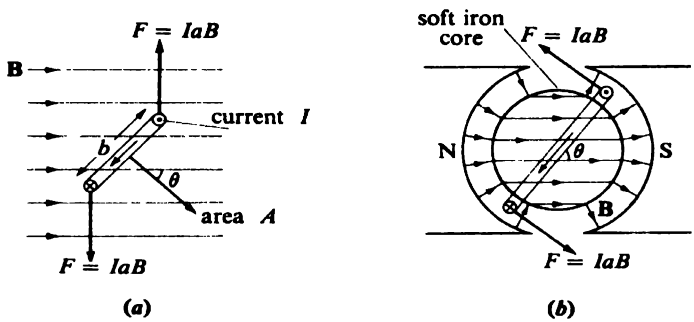


Fig. 8.12. The couple on a plane coil in (a) uniform and (b) radial magnetic fields.

and if the net introduced is in the plane of the circuit each elementary mesh experiences a couple $I\,dSB \sin\theta$. Because I, B and θ are the same for all the meshes the total couple is $IAB \sin\theta$ or, for N turns,

$$T_\theta = NIAB \sin\theta \quad \text{to decrease } \theta \tag{8.18}$$

or

$$\mathbf{T} = NI\mathbf{A}\times\mathbf{B}$$

A more general formula, applicable in non-uniform fields, can be obtained through the expression for the potential energy of a dipole in a field $U = -\mathbf{m}\cdot\mathbf{B}$ (equation (7.15)). For a small current loop of moment $I\,d\mathbf{S}$, $U = -I\,d\mathbf{S}\cdot\mathbf{B}$ or $-I\,d\Phi$ where $d\Phi$ is the external

magnetic flux linking the loop, counted positive when in the direction of **m** or of the self-flux. For a large circuit this becomes

$$U = -I\Phi \tag{8.19}$$

and from this we can calculate forces and couples assuming (as does the derivation of (7.15)) that the currents are invariant: for, since U is the work done by external forces or couples in establishing the circuit in its position in the field, the force in any direction **s** is given by $F_s\,\mathrm{d}s = -\mathrm{d}U$ and the couple about any axis by $T_\theta\,\mathrm{d}\theta = -\mathrm{d}U$. Hence,

$$F_s = -\left(\frac{\partial U}{\partial s}\right)_I = I\left(\frac{\partial \Phi}{\partial s}\right)_I; \qquad T_\theta = -\left(\frac{\partial U}{\partial \theta}\right)_I = I\left(\frac{\partial \Phi}{\partial \theta}\right)_I \tag{8.20}$$

Application of this to the plane coil of Fig. 8.12a for which the flux linkage is $NAB\cos\theta$ confirms (8.18) above. For the coil of Fig. 8.12b in a radial field, the flux linkage is $NAB\theta$ and

$$T_\theta = NABI \tag{8.21}$$

Equations (8.20) show that a circuit will be in equilibrium ($\mathbf{F}=\mathbf{T}=0$) when Φ has a stationary value: the equilibrium will be stable if a small displacement brings into play restoring forces or couples, i.e. if the flux decreases away from the point of equilibrium. Thus *for stable equilibrium a circuit is embracing maximum flux*, and in any other position the forces or couples will act in such directions as to cause the circuit to increase the flux linkage.

Although U is the mechanical potential energy of a circuit in a field we cannot identify it with the magnetic energy: we shall see in the next chapter that to maintain a current constant when it moves in a magnetic field requires an input of energy from e.m.f.s and the complete energy balance must include this.

8.5 Currents in Magnetic Fields 2: The Current Element Formula

In Fig. 8.13 a current element as part of a complete circuit produces the field at P given by (8.15) in the direction shown. At P a small magnetic dipole of moment **m** is placed with its axis pointing at the element and, because **B** varies with r, this dipole experiences a force given by $m\,\mathrm{d}B/\mathrm{d}r$ (from (7.14) or treat the dipole as two poles) in the direction of **B** itself. Thus the element exerts a force on the dipole

$$\begin{aligned} \mathrm{d}F &= (m\mu_0 2I\,\mathrm{d}l\sin\theta)/4\pi r^3 \\ &= I\,\mathrm{d}l\sin\theta\,\frac{\mu_0}{4\pi}\frac{2m}{r^3} \end{aligned} \tag{8.22}$$

An equal and opposite force is exerted on the element by the dipole. The last part of (8.22) is in fact just the magnitude of **B** *at the element* produced by the dipole and since we are adopting in general the assumption that there is no distinction between one magnetic field

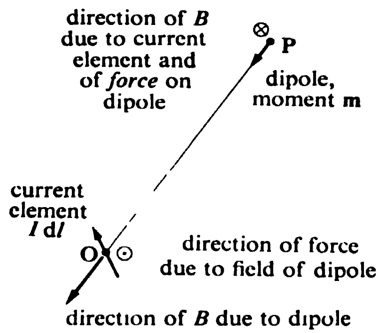


Fig. 8.13. Derivation of the force on a current element in a magnetic flux density **B.**

and another, we should find that the force on an element in a field of flux density B is given by

$$dF = BI\,dl\sin\theta \tag{8.23}$$

(at right angles to I and B such that I, B and dF form a right-handed system) or, more concisely in vector form,

$$\mathbf{dF} = I\,\mathbf{dl}\times\mathbf{B} \tag{8.24}$$

The directions of the various quantities are sometimes remembered by Fleming's left-hand rule as long as $\theta = \frac{1}{2}\pi$, but the vector form is at once more general and more concise.

Unlike the current element formula (8.15) or (8.16), this one can be checked experimentally because it is possible to measure the force on *part* of a circuit. Ampère showed that the force was perpendicular to an element by mounting one on mercury cups as in Fig. 8.14a and allowing it only one degree of freedom, along its own length: no magnets or currents brought up to it caused it to move. The balance of Fig. 8.14b uses a shape of circuit such that only the force on the portion XY produces any moment about the fulcrum, the upper end being in a region of negligible field. The balance can be used either to check (8.23) or to measure B (Thomas, Driscoll and Hipple, 1950, used the method in precise measurement of the gyromagnetic ratio of the proton—see section 11.6).

Couple on Plane Coils in Magnetic Fields. If the coils of Figs. 8.12a and 8.12b are rectangular and pivoted about axes through the centre as shown, the turning forces will be those acting on the sides perpendicular to the plane of the diagram and will be $NIaB$ by (8.23) where a is the length of the side. The turning couples are

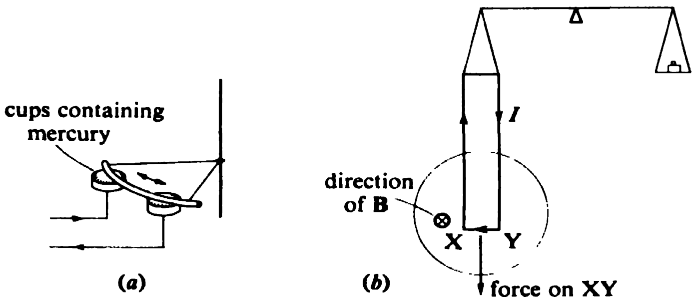


Fig. 8.14. Forces on current elements (a) Ampère's experiment; (b) the balance of Thomas, Driscoll and Hipple.

thus $NIAB \sin\theta$ for Fig. 8.12a and $NIAB$ for Fig. 8.12b, A being the area of the coils. These confirm (8.18) and (8.21) for rectangular coils: any plane shape can be divided into elementary rectangles with identical I, B and θ's and the couples summed to give the general form of (8.18) and (8.21). These couples are those acting in electric motors and in galvanometers (section 16.5).

8.6 Forces between Currents

It is possible to combine (8.16) and (8.24) into a single formula giving the force between current elements as

$$d^2\mathbf{F} = \frac{\mu_0 I_1 I_2 \, d\mathbf{l}_1 \times (d\mathbf{l}_2 \times \mathbf{r})}{4\pi r^3} \tag{8.25}$$

and regard this as the basic magnetic law of force. With this approach $\mathbf{B}$ would be immediately defined by (8.16) and equation (8.24) would follow—this once again corresponds to the splitting of Coulomb's law (cf. section 7.7).

Although the forces on $d\mathbf{l}_1$ and $d\mathbf{l}_2$ given by (8.25) are not equal and opposite, we realize that the formula has no physical meaning in that an element of a circuit carrying a steady current cannot produce its force separate from the rest of the circuit: when (8.25)

is integrated round the complete circuits the forces obtained are then equal and opposite.

To find the mutual forces and the couples between two complete current-carrying circuits we first find **B** due to one circuit and then calculate the force or couple on the other in **B**. Equation (8.25), integrated round both circuits by using (8.16) and (8.24), represents one way of carrying this out, while other ways are illustrated by the two examples immediately following. Alternatively, the mutual forces and couples can be calculated using formulae to be developed in section 9.11.

Force between Two Infinite Straight Parallel Currents. Two infinite wires carrying currents I_1 and I_2 are parallel and a constant distance r apart (Fig. 8.15). The flux density at every point on I_2

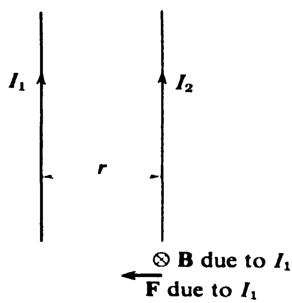


Fig. 8.15. The force between two infinite parallel currents.

due to I_1 is, by (8.11), $\mu_0 I_1/2\pi r$ into the diagram and the force on a length l of I_2 is therefore, by (8.23),

$$F = \frac{\mu_0 I_1 I_2 l}{2\pi r} \tag{8.26}$$

towards I_1. The reversal of I_2 would reverse the force as the experiment of section 1.2 showed.

The whole of chapters 7 and 8 could have been developed to this point using a wholly arbitrary unit of current by retaining μ_0, first introduced in equation (7.6), as a constant whose value depends on our choice. In the MKSA system the ampere is defined as in section 1.4 and this, by (8.26), fixes μ_0 in that system at $4\pi \times 10^{-7}$ henries/m (for the henry see section 9.5).

The CGS e.m.u. system, although originally based on the unit

pole, can just as well be based on a unit of current defined similarly to the ampere but with a force of 2 dyn/cm length between the wires when 1 cm apart. From (8.26) the value of μ_0 in this system will then be 4π which gives the law of force between poles as P_1P_2/r^2 from (7.13). For the value of μ_0 in CGS e.s.u. see section 16.7.

Equation (8.26) will also give an approximate value for the force between any two coils whose distance apart is much smaller than any radius of curvature of the wires.

Application of Dipole Formulae. When one or both of two current-carrying circuits are small compared with their distance apart, dipole formulae may be used. For example, in Fig. 8.16, the

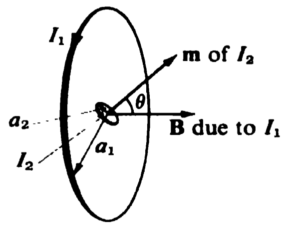


Fig. 8.16. A small circular coil at the centre of a large one.

small coil at the centre can be treated as a dipole of moment $I_2\pi a_2{}^2$ at θ to a field $\mu_0 I_1/2a_1$, producing a couple

$$T_\theta = \frac{\mu_0 I_1 I_2 \pi a_2{}^2 \sin\theta}{2a_1} \tag{8.27}$$

For forces, (7.14) may be used on the understanding that we shall justify it for current loops later (section 8.8).

8.7 Moving Charges

A current element $I\,\mathrm{d}l$ consists of charges in motion over a small length of a circuit. By (1.15), $I = nQvA$ where A is the cross-section of the element, n is the number of charges per unit volume each of magnitude Q, and v is their mean velocity. Thus

$$I\,\mathrm{d}\mathbf{l} = nQ\mathbf{v}A\,\mathrm{d}l = NQ\mathbf{v} \tag{8.28}$$

where N is the total number of charges moving in the element. N may be large even if $\mathrm{d}l$ is small compared with other distances.

Flux Density due to a Moving Charge. It would appear from (8.15) and (8.16) that the flux density due to N moving charges

would be $(\mu_0 NQv \sin\theta)/4\pi r^2$ using (8.28) and that due to one charge only

$$B = (\mu_0 Qv \sin\theta)/4\pi r^2 \tag{8.29}$$

(in a direction given by the right-hand screw rule) or, vectorially,

$$\mathbf{B} = \frac{\mu_0 Q\mathbf{v}\times\mathbf{r}}{4\pi r^3} \tag{8.30}$$

(Fig. 8.17a). This formula, like (8.15), cannot be checked directly and must be used with caution particularly since the original

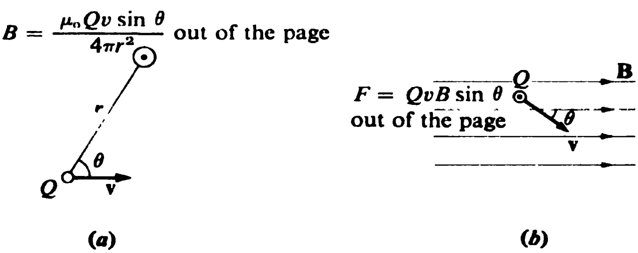


Fig. 8.17. (a) **B** *due to a moving charge; (b) the force on a moving charge in* **B**.

current element formula was derived for steady currents (see also problem 15.3). It is shown in more advanced texts that (8.30) is only applicable for velocities small compared with that of light.

Force on a Moving Charge in a Magnetic Field. From (8.23) and (8.28), the force on N charges in a magnetic flux density B would be $NQvB \sin\theta$ and on a single charge

$$F = QvB \sin\theta \tag{8.31}$$

at right angles to v and B such that v, B and F form a right-handed system. Vectorially,

$$\mathbf{F} = Q\mathbf{v}\times\mathbf{B} \tag{8.32}$$

(Fig. 8.17b). This formula can be accepted with much more confidence because of our greater faith in the corresponding current element formula (8.24) and because it has consequences, explored in chapter 11, which can be checked experimentally.

Some authors use (8.32) as a definition of **B**: note that **F** is a force on a moving charge over and above any electrostatic force which might exist.

8.8 General Laws for Steady Magnetic Fields *in Vacuo*

In section 4.9, the general laws applicable to steady electrostatic fields were summarized by quoting the surface integral of **E** over a closed surface (Gauss's theorem) and the line integral round a closed path (e.m.f. = 0) which together gave the sources and vortices of **E**. We now see that the laws for steady magnetic fields can be similarly summarized by

$$\oint_C \mathbf{B}\cdot d\mathbf{s} = \mu_0 I \tag{8.10}$$

the circuital theorem and by

$$\oint\!\!\oint_S \mathbf{B}\cdot d\mathbf{S} = 0 \tag{8.33}$$

which is Gauss's theorem. Since both these integrals were zero for paths *in vacuo* in the field of poles only, they will not be affected by the presence of poles although they do not necessarily apply inside magnetic materials (chapter 14). We may say that currents are vortices of **B** but not sources: lines of **B** do not end or originate on currents.

It is beyond the scope of this book to justify Gauss's theorem rigorously but the following argument is satisfactory. Figure 8.18 shows a current element within a surface S. For a given r and θ, B is constant round a path such as C and a small tube with C as axis cuts S in two places so that at one intersection B is inwards and at

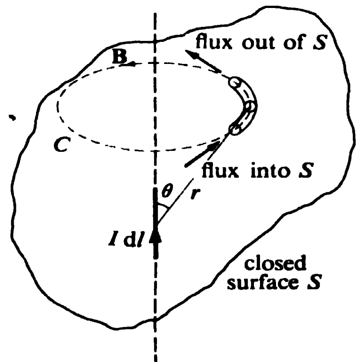


Fig. 8.18. Derivation of Gauss's theorem for steady currents.

the other it is outwards. Because B is the same, the flux across the two intersections is equal whatever the angle at which S is cut by the tube and the resultant outward flux is therefore zero. Any system of currents can be split into similar elements and (8.33) thus justified.

Differential Forms. By methods similar to those used to obtain (4.3) and (4.6), the following equations can be derived from (8.10):

$$\begin{aligned} \partial B_z/\partial y - \partial B_y/\partial z &= \mu_0 J_x \\ \partial B_x/\partial z - \partial B_z/\partial x &= \mu_0 J_y \\ \partial B_y/\partial x - \partial B_x/\partial y &= \mu_0 J_z \end{aligned} \tag{8.34}$$

and the following from (8.33):

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \tag{8.35}$$

J_x, J_y and J_z are the current density components.

The more advanced reader is referred to problem 8.16 for an indication of the way in which equations (8.34) and (8.35) can be expressed more concisely by using the operators *curl* and *divergence*.

Justification of Dipole Force Formula. The formula giving the force on a dipole in a non-uniform field, (7.14), originated from a proof assuming the existence of two poles. We have not justified it for a current loop and we shall do so now only by considering the special case of a dipole with a moment along the x-axis in a field which is non-uniform but has cylindrical symmetry about the x-axis. This case brings out clearly the reason for the force.

In Fig. 8.19, we see a small current loop with its moment $m = IA = I\pi a^2$ lying along B_x. We may well ask why any force should arise since the loop has no extension in the x-direction as has a two-pole dipole. The reason lies in equation (8.35) from which it is clear that a non-uniformity along x must be accompanied by non-uniformity in some other direction and since the loop *has* extension

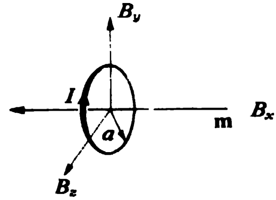


***Fig. 8.19.** Force on a current-loop dipole in a non-uniform magnetic field.*

in other directions we must look to B_y, B_z, etc., for the force. With cylindrical symmetry the component of B perpendicular to the x-axis will have the same value at all points on the loop, the same as B_y or B_z. If B at the centre is along the x-axis, then both B_y and B_z at the circumference of the loop have the value $a\,\partial B_y/\partial y$ or $a\,\partial B_z/\partial z$ which are therefore equal and, by (8.35) are equal to $\frac{1}{2}a\,\partial B_x/\partial x$. This same B acts on I in a wire of length $2\pi a$. Thus the total force *along* x is $I\,2\pi a\,\frac{1}{2}a\,\partial B_x/\partial x$ which is $m\,\partial B_x/\partial x$ as for a permanent dipole.

8.9 Summary of Chapter 8

Four methods have been developed for calculating **B** due to currents:

1. The use of dipole formulae, applicable to small circuits.

2. The use of an equivalent magnetic shell if the solid angle subtended by the circuit at the point can be calculated. If it can,

$$V_m = I\Omega/4\pi \quad \text{and} \quad B_x = -\mu_0\,\partial V_m/\partial x, \text{ etc.} \tag{8.3}$$

3. The use of the circuital theorem in cases of high symmetry:

$$\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0 I \tag{8.10}$$

4. The use of the current element formula:

$$d\mathbf{B} = \mu_0 I\, d\mathbf{l}\times\mathbf{r}/4\pi r^3 \tag{8.16}$$

Three methods have been developed for calculating forces on currents in magnetic fields:

1. The use of dipole formulae, applicable to small circuits.
2. The use of the equivalent magnetic shell leading to

$$F_s = I(\partial\Phi/\partial s)_I; \qquad T_\theta = I(\partial\Phi/\partial\theta)_I \tag{8.20}$$

3. The use of the current element formula

$$d\mathbf{F} = I\, d\mathbf{l}\times\mathbf{B} \tag{8.24}$$

From the current element formulae have come expressions for the flux density due to and force on a moving charge:

$$\mathbf{B} = \mu_0 Q\mathbf{v}\times\mathbf{r}/4\pi r^3 \quad \text{and} \quad \mathbf{F} = Q\mathbf{v}\times\mathbf{B} \quad \text{(8.30) and (8.32)}$$

the first of which must be used with circumspection.

Finally we have summarized the properties of steady magnetic fields by the circuital theorem and Gauss's theorem:

$$\oint \mathbf{B}\cdot d\mathbf{s} = I \quad \text{and} \quad \oiint \mathbf{B}\cdot d\mathbf{S} = 0 \quad \text{(8.10) and (8.33)}$$

PROBLEMS

SECTION 8.1

8.1 Confirm that the distance apart of a pair of Helmholtz coils is equal to their radius.

8.2 A uniform magnetic field is required over a small region of space. Enumerate three ways of producing such a field and compare them.

8.3 Expose the fallacy in the following. The magnetic scalar potential at a point outside a small current-carrying loop, but in the same plane as the loop, is zero because the solid angle is zero. Hence the magnetic flux density is zero and a current-loop dipole produces no broadside field.

SECTION 8.2

8.4 A steady current I flows in one direction in the solid inner conductor, radius a, of a coaxial cable, and in the opposite direction in the outer conductor whose inner radius is b and outer radius c. Find the magnetic flux densities at various distances from the axis.

8.5 Express the flux density 1 m from an infinite straight wire carrying 1 A in (*a*) Wb/m^2 (*b*) gauss, and hence find a conversion factor between the two.

SECTION 8.3

8.6 Find the magnetic flux density (*a*) at the centre of a square coil of side a carrying a current I, (*b*) a distance x along the perpendicular from the centre of the square coil in (*a*), (*c*) at the centre of a regular polygon of n sides inscribed in a circle of radius r and carrying a current I.

8.7 A magnetic shell of strength I is in the shape of a cap of a sphere of radius a centre P, the cap subtending a semi-vertical angle θ at P. Find the magnetic flux density at P and show that it tends to zero as a tends to infinity

8.8 An infinite plane current sheet has a uniform surface current density J_s (see equation (1.14)). Find the magnetic flux density outside the sheet.

SECTION 8.5

8.9 A uniform straight rod of mass m per unit length is hung vertically so that it can turn about its upper end, and a current I is sent through it. Find how much it is deflected from the vertical in a place where the earth's horizontal magnetic flux density is B_0.

8.10 Find the force on a current-carrying wire in the shape of an arc of a circle placed in a plane perpendicular to a uniform magnetic field. Show that the same force is exerted if the same current flows in the corresponding chord of the circle and generalize the result.

8.11 A light circular flexible loop of radius a carries a current I and is placed in a plane perpendicular to a uniform magnetic flux density $\boldsymbol{B}$. Find the tension in the loop.

SECTION 8.6

8.12 Two flat coils each of 10 turns have mean radii of 20 cm. Find an approximate value for the force between them if they are coaxial, 1 cm apart and each carries 5 A in the same sense.

8.13 Two flat coils each of 10 turns have mean radii of 20 cm and 2 cm. Find an approximate value for the force between them if they are coaxial, 10 cm apart and if each carries 5 A in the same sense.

SECTION 8.7

8.14 A wire is bent into a circle of radius a and a total charge Q is distributed uniformly around it. Find the magnetic flux density at the centre if the wire rotates with an angular velocity ω about the axis of the circle.

8.15 An insulating circular disc of radius a is sprayed with charge to a uniform surface density σ. If the disc rotates about its axis with an angular velocity ω, what is the magnetic flux density at its centre?

SECTION 8.8

*8.16 If the symbol $\boldsymbol{\nabla}$ (del) stands for the differential operator $\mathbf{i}(\partial/\partial x)+\mathbf{j}(\partial/\partial y)+\mathbf{k}(\partial/\partial z)$, show that (8.34) can be written $\boldsymbol{\nabla}\times\mathbf{B}=\mu_0\mathbf{J}$, and (8.35) as $\boldsymbol{\nabla}\cdot\mathbf{B}=0$. ($\boldsymbol{\nabla}\times\mathbf{B}$ is also written, and called, curl $\mathbf{B}$; $\boldsymbol{\nabla}\cdot\mathbf{B}$ is also written div $\mathbf{B}$ and called the divergence of $\mathbf{B}$.)

*8.17 Show that the relation between $\mathbf{E}$ and V (section 3.6) can be written $\mathbf{E}=-\boldsymbol{\nabla}V$. ($\boldsymbol{\nabla}V$ is also written **grad** V and called the gradient of V.)

8.18 Show that $\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\equiv\nabla^2=\partial^2/\partial x^2+\partial^2/\partial y^2+\partial^2/\partial z^2$ as in section 4.2.

CHAPTER 9

ELECTROMAGNETIC INDUCTION, INDUCTANCE AND MAGNETIC ENERGY

The only variations with time so far encountered have been the acceleration of charges by electric fields in chapter 3 and the charge and discharge of a condenser in chapter 6. In both cases the laws established for steady conditions were assumed to apply and the extension justified, at any rate approximately, by experiment. In this chapter we turn to a proper investigation of time variations in *magnetic* fields: the corresponding problem in electric fields is left until chapter 15.

9.1 Electromagnetic Induction: Experimental Basis

Faraday's discovery of the first* induced current on 29th August 1831 began six months of brilliant experiment in which he established most of the qualitative laws of electromagnetic induction. In this section we shall summarize his results and those of his contemporaries in terms of the concepts met in previous chapters, remembering that these were only then being painfully developed: not the least of Faraday's achievements was the expression of his results in terms of lines of force and flux, concepts which later proved so fruitful.

Faraday's experiments of 1831–32 included the following:

1. Two coils are arranged so that if a steady current flows in one some of its magnetic flux links the other. If the current in the first coil changes, a current is induced in the second: this is known as *mutual induction*. (Faraday used two helices wound on a common iron core and later dispensed with the core.)
2. A coil is arranged to link some of the magnetic flux from a source S of magnetic field which may be either a magnet or a current. If relative motion occurs between the coil and S such

* Strictly the first induced current *recognized as such*. Arago in 1824 had observed the motion of a suspended magnet induced by a rotating copper disc beneath it (Arago's disc) but could not account for it.

that the flux linking the coil changes, a current is induced in it. (Faraday plunged a bar magnet into a solenoid; moved a coil near and between the poles of a magnet; made one current-carrying coil approach another.)
3. Part of a conducting circuit moving and thereby cutting magnetic flux has a current induced in it. (Faraday's disc, illustrated in Fig. 9.1, was the first continuous generator: he also passed a straight wire between the poles of a magnet.)

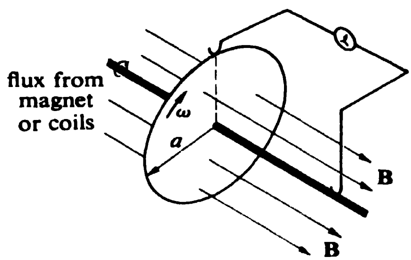


Fig. 9.1. Faraday's disc, a homopolar generator.

4. The induced currents produced under the same conditions are proportional to the conductance of the circuit, i.e. given changes produce a definite e.m.f. rather than a definite current.
5. Greater rates of change produce larger e.m.f.s.

That an e.m.f. is induced in a circuit because of a change in its own current (*self-induction*) was first discovered by Joseph Henry (1797–1878) in 1832 and independently investigated by Faraday in 1834.

Faraday's laws of electromagnetic induction may thus be summarized as follows. An e.m.f. is induced (*a*) in a rigid stationary circuit across which there is a changing magnetic flux, (*b*) in a rigid circuit moving in a steady magnetic field in such a way that the flux across it changes and (*c*) in *part* of a circuit which *cuts* any magnetic flux. Induced e.m.f.s of type (*a*) we shall call *changing-field e.m.f.s*, and those of types (*b*) and (*c*) *motional e.m.f.s*. Phenomena encountered in practice can usually be placed in one category or the other, although it is clearly possible both for a field to be changing in time and for a circuit to be simultaneously moving through it.

Faraday's original descriptions of the *direction* of the induced e.m.f.s were confused and Emil Lenz in 1834 gave the first clear statement. Lenz's law is now usually given in the form: whenever

a change produces an induced current, the direction of current flow is such as to produce effects opposing the change. Some examples of its operation are given in Fig. 9.2, which should be carefully studied.

There is no explicit reference to the *magnitude* of the induced e.m.f. in these early researches and it was not until 1845 that Neumann expressly assumed the proportionality of e.m.f. and rate of change of flux linkage or rate of flux cutting. Faraday showed in

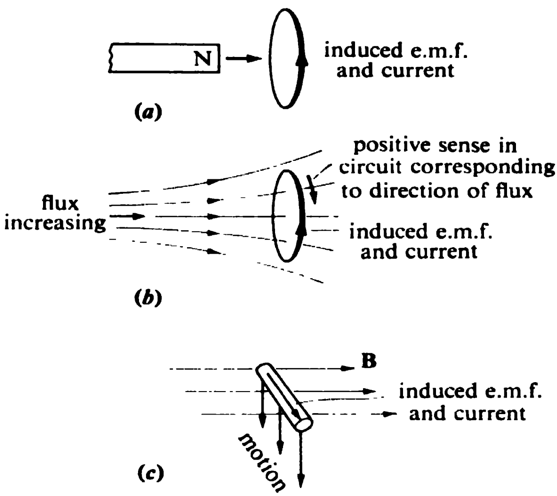


Fig. 9.2. ***Operation of Lenz's law. The direction of induced currents (a) opposes motion of*** N ***pole and produces flux in opposition to increase of flux, (b) produces flux in opposition to increase of flux, (c) produces an upward force opposing motion.***

1851–52 that motional e.m.f.s were proportional to the rate of flux cut but the law for stationary circuits seems merely to have been assumed.

9.2 Electromagnetic Induction: Deductions from Previous Results

We now derive certain results from laws established in previous chapters together with some assumptions.

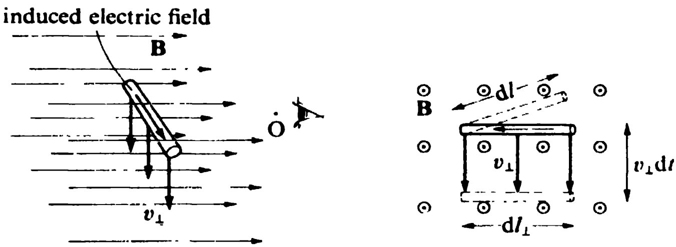
Motional E.m.f.s: Current Element Formula. Any charge Q moving with a velocity $\mathbf{v}$ in a magnetic flux density $\mathbf{B}$ experiences a force

$$\mathbf{F} = Q\mathbf{v}\times\mathbf{B} \qquad (8.32)=(9.1)$$

When a straight conducting wire moves in a magnetic field this force is exerted on all the charges in it, so that by definition (3.1) they will experience an electric field given by

$$\mathbf{E} = \mathbf{F}/Q = \mathbf{v}\times\mathbf{B} \tag{9.2}$$

We remember that (3.1) was restricted to cases where Q was stationary so that it is *from the point of view of the conductor* that $\mathbf{E}$ exists. Figure 9.3 shows the relation between directions for the



(*a*) (*b*)

Fig. 9.3. ***(a) Motional e.m.f. in an element of wire; (b) view of (a) from O.***

special case of a wire at right angles to $\mathbf{B}$ moving with a velocity perpendicular both to the wire and to $\mathbf{B}$. The induced electric field will cause a current if the wire is connected to a conducting circuit and the direction of the current is remembered in elementary physics by Fleming's right-hand rule. Once the vector product has been mastered, however, (9.2) is more concise and more general and shows that only the component of velocity perpendicular to $\mathbf{B}$ ($v_\perp$) contributes to $\mathbf{E}$. Thus (9.2) could be expressed as

$$E = v_\perp B \quad \text{in magnitude} \tag{9.3}$$

The e.m.f. across the ends of an element $\mathbf{dl}$ will, from (9.2), be

$$d\mathscr{E} = \mathbf{E}\cdot\mathbf{dl} = (\mathbf{v}\times\mathbf{B})\cdot\mathbf{dl} \tag{9.4}$$

but this can be simplified by realizing that the scalar product means that $\mathbf{dl}$ should be resolved in the direction of $\mathbf{E}$, i.e. that only the component of $\mathbf{dl}$ perpendicular to $v_\perp$ and $\mathbf{B}$, $dl_\perp$, is significant. Hence

$$d\mathscr{E} = v_\perp B\, dl_\perp \quad \text{in magnitude} \tag{9.5}$$

which can be regarded as a current element formula for induced e.m.f.s and can be summed round a circuit to give a total e.m.f.

For a complete circuit therefore

$$\mathscr{E} = \oint(\mathbf{v}\times\mathbf{B})\cdot\mathrm{d}\mathbf{l} = \oint v_{\perp}B\,\mathrm{d}l_{\perp} \tag{9.6}$$

although the form (9.5) is the most useful.

Motional E.m.f.s: Flux-cutting Law. Figure 9.3b shows that $v_{\perp}\,\mathrm{d}t\;\mathrm{d}l_{\perp}$ is the area traced out by the element in time $\mathrm{d}t$ so that the flux cut or crossed by it is $v_{\perp}B\,\mathrm{d}l_{\perp}\,\mathrm{d}t$ or $\mathrm{d}\mathscr{E}\,\mathrm{d}t$ by (9.5). Thus the e.m.f. induced in the element is equal to the flux cut by it per unit time, and summing round a complete circuit

$$\mathscr{E} = \frac{\mathrm{d}\Phi}{\mathrm{d}t} \qquad (\Phi \text{ is flux } \textit{cut}) \tag{9.7}$$

where Φ is to be obtained by summing the fluxes cut by the elements *with due regard to sign.* The positive sense of a circuit is related to the flux linking it by the right-hand screw rule (Fig. 9.4a), and the flux cut by an element is positive by (9.7) if the induced e.m.f. acts in a positive direction round the circuit. The flux-cutting law is most useful when only part of a circuit is moving and the direction of the e.m.f. is then most easily obtained from that of $(\mathbf{v}\times\mathbf{B})$.

Motional E.m.f.s: Flux-linking Law. The flux cut by a moving rigid circuit in time $\mathrm{d}t$ is the flux which crosses the curved surface of the volume traced out by it (Fig. 9.4b): let this be $\mathrm{d}\Phi_{\mathrm{cut}}$ using the sign convention above. The fluxes *linked* by the same circuit in the initial and final positions are the fluxes across the ends of the volume traced out, and if these fluxes are Φ_{i} and Φ_{f} $(=\Phi_{\mathrm{i}}+\mathrm{d}\Phi_{\mathrm{link}})$ respectively, the change in flux linkage is $\mathrm{d}\Phi_{\mathrm{link}}$. Since there are no sources of $\mathbf{B}$ within the volume, Gauss's theorem shows that the total outward flux over the closed surface must be zero. It follows that

$$\mathrm{d}\Phi_{\mathrm{link}}+\mathrm{d}\Phi_{\mathrm{cut}} = 0$$

in all cases. (In Fig. 9.4b, the outward flux is $+\mathrm{d}\Phi_{\mathrm{link}}-(-\mathrm{d}\Phi_{\mathrm{cut}})$ while in Fig. 9.4c it is $-\mathrm{d}\Phi_{\mathrm{link}}-\mathrm{d}\Phi_{\mathrm{cut}}$.)

Hence

$$\mathscr{E} = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} \qquad (\Phi \text{ is flux } \textit{linked}) \tag{9.8}$$

The negative sign means that if the motion is one which increases the flux linkage, the e.m.f. is in a negative sense round the circuit and therefore produces currents which oppose the motion. Note that (9.6), (9.7) and (9.8) are *alternative* formulae for $\mathscr{E}$.

Changing-field E.m.f.s: Flux-linking Law. Although (9.8) has been derived for a circuit moving in a field, we make the important assumption (based on experience) that it is only *relative* motion between circuit and source of field which matters. Hence if $d\Phi/dt$ is the rate of change of flux linkage across a stationary circuit due to

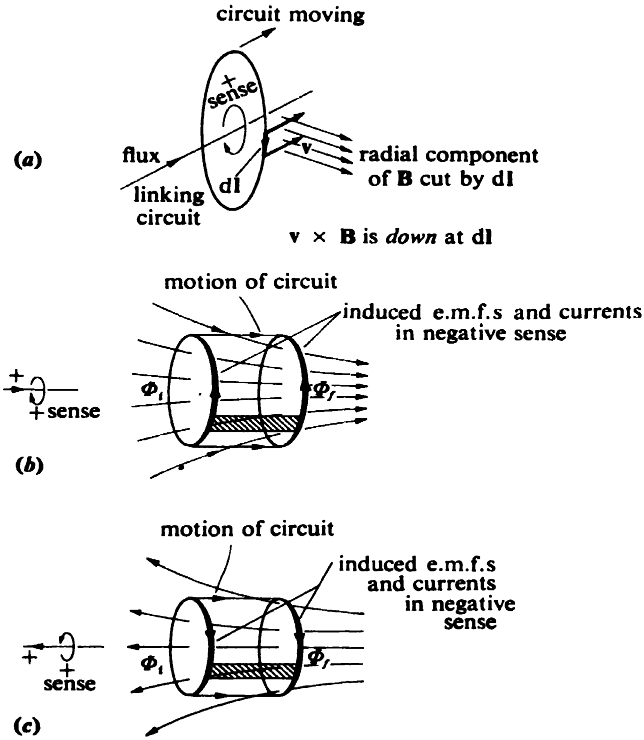


Fig. 9.4. Flux cutting and flux linking. (a) Flux cut by dl *is positive because* $\mathbf{v}\times\mathbf{B}$ *is in the positive sense of the circuit; (b) and (c) movement of the circuit to increase positive flux linkage, showing that* $d\Phi_{cut}+d\Phi_{link}=0$.

a moving source, equation (9.8) should apply equally well. If we also assume that no distinction can be drawn between fields which change because of a moving source and those which change because of, say, mutual induction, we again expect (9.8) to apply.

Thus for stationary circuits the law of electromagnetic induction takes the form

$$\mathscr{E} = -\frac{\partial \Phi}{\partial t} \tag{9.9}$$

the partial derivative occurring because only changes in Φ with time are important, not changes with position. The assumptions used above in deriving this law are justified by its consequences, particularly in A.C. circuits (chapter 10). We shall refer to (9.9) as *Neumann's law*.

Remembering the meaning of e.m.f. from chapter 6, we see that round any path C which encloses a changing magnetic flux we expect an electric field to follow closed lines just as **B** does in the space round a current. Such an electric field would not produce a path-independent potential difference any more than **B** did. This aspect will be pursued further in sections 11.4 and 15.1. When a conducting wire coincides with C a current is produced and the e.m.f. is non-localized (section 6.1). If, however, only a part of a circuit moves in a magnetic field, a localized e.m.f. is produced.

Units. From (9.7), etc., the unit of flux is the volt-second which we introduced in section 7.3 as the *weber*. The CGS e.m.u. is the *maxwell*.

9.3 Induced Currents and Charges

The induced currents in masses of metal (e.g. in generators, motors and transformers) are known as *eddy currents* and are generally undesirable because of the RI^2 heat losses. In addition, if the metal is moving they flow in such directions as to produce forces opposing the motion. Eddy currents are minimized by laminating the metal, using high resistance materials or dispersing the metal in a non-conducting matrix, but their damping effect can be used (e.g. in galvanometers, section 16.6, and in electricity supply watt-hour meters where the Arago's disc principle is used to control the speed of the motor).

In a conducting wire circuit of total resistance R, the induced current will be

$$I = -\frac{1}{R}\frac{\mathrm{d}\Phi}{\mathrm{d}t} \tag{9.10}$$

while the charge flowing in a circuit in which the flux linkage

changes from Φ_i to Φ_f in a time from t_i to t_f will be

$$Q = \int_{t_i}^{t_f} I \, dt = -\frac{1}{R}\int_{\Phi_i}^{\Phi_f} d\Phi$$

$$Q = \frac{\delta\Phi}{R} \tag{9.11}$$

where $\delta\Phi$ is the change in flux linkage.

Search Coil. An important method for the measurement of magnetic flux and flux density uses a small coil of many turns, N, each of mean area A, placed so that the required B crosses it normally (Fig. 9.5). The flux linkage is then NAB and a flux *change* of NAB or $2NAB$ is produced by removing the coil to a

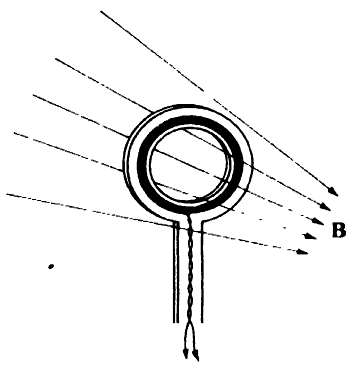


Fig. 9.5. A search coil.

place where $B=0$ or by reversing the direction of B, whichever is the more convenient. The flux change is measured directly by a fluxmeter (section 16.6) or, if produced quickly enough, the charge given by (9.11) is measured by a ballistic galvanometer. For rough measurements of B it may be sufficient to calculate the constant NA from the geometry of the search coil, but it is usual to calibrate any arrangement with a standard B obtained from a solenoid or with a standard flux change from a mutual inductance.

The method gives a mean value for B over the area of the coil, although search coils have been designed which give accurately the magnitude of B at the central point (Brown and Sweer's flux ball (1922); Herzog and Tischler (1923)).

9.4 Motional E.m.f.s: Generators and Motors

Apart from eddy current damping and the search coil, the chief applications of motional e.m.f.s lie in the generation of electric power and the associated subject of electric motors. These are highly technical matters and we shall only touch on them.

All generators and motors consist of a *rotor* and a *stator*. One of these, usually the stator, produces a magnetic field by coils carrying a current and wound on pole pieces. The other, usually the rotor and in that case known as the armature, carries a current either induced by the relative motion (in a generator) or producing relative motion (in a motor). This current, whether generated or injected, must be connected to stationary input or output terminals through carbon brushes in contact with slip rings or commutator segments which revolve with the rotor. Large A.C. generators reverse the general rule and have rotating field coils and pole pieces because the brushes then have to carry only the relatively small field currents and the heavy duty wiring is supported in the stator. All generators, when supplied with current from an outside source, become motors although sometimes small modifications are needed.

We classify machines into homopolar, heteropolar and induction as in table 9.1.

Homopolar Machines. The Faraday disc of Fig. 9.1 can be modified to form a homopolar generator, the field usually being produced by current-carrying coils rather than by magnets as in the original apparatus. The usual arrangement, as in the Lorenz disc method (section 16.2) has a **B** which is symmetrical about the axis of the disc but which varies radially. In this case, if Φ is the total flux crossing the disc face, a radius of length a cuts a flux $\Phi \, d\theta/2\pi$ in a time dt. The induced e.m.f. is thus $\Phi\omega/2\pi$ where ω is the angular velocity, or

$$\mathscr{E} = n\Phi \qquad (9.12)$$

where n is the number of revolutions per unit time. Such a generator has been recently used to produce currents for generating steady magnetic fields.

Heteropolar Machines. A coil of N turns each of mean area A rotates in a uniform magnetic flux density B (Fig. 9.6a) at an angular velocity ω. At any instant when the normal to the plane of the coil makes an angle θ with B, the flux linkage is $NAB \cos \theta$. The induced e.m.f. is thus

$$\mathscr{E} = NAB\omega \sin \omega t = \Phi_{max}\omega \sin \omega t \qquad (9.13)$$

Table 9.1

COMMON GENERATORS AND MOTORS

Type	*Rotor*	*Stator*	*Production of field in generator by*	*Brushes distribute current by*	*Generator*	*Motor*
I. Homopolar	Metal disc	Perm. mag. or Pole pieces and field windings	Perm. mag. or Separately excited	Slip rings	Faraday disc (D.C.) Lorenz disc, Homopolar generator	Barlow's wheel (D.C.)
II. Heteropolar	Drum armature with multiple windings	Pole pieces and field windings	Self-excited	Commutators	D.C. generators	D.C. motors, small A.C. motors
			(*a*) series wound		little used	large starting torque
			(*b*) shunt or compound wound		output varies little with current drawn	torque varies little with load
			Separately excited			
III. Heteropolar	Drum armature—one coil per phase	Pole pieces and field windings	Separately excited	Slip rings	Small A.C. generators	A.C. synchronous motors
IV. Heteropolar	Pole pieces and field windings	Stator windings—one set per phase	Separately excited	Slip rings (field current)	Large A.C. generators	A.C. synchronous motors
V. Induction	'Squirrel cage' of conductors	Field windings—one set per phase	—	—	—	A.C. induction motor

in magnitude, where Φ_{max} is the maximum flux linked by the coil. The current taken off by slip rings is alternating (A.C.) and Fig. 9.6b shows the form of a 3-phase output which could be obtained by winding three coils at 120° to each other. If the field is produced

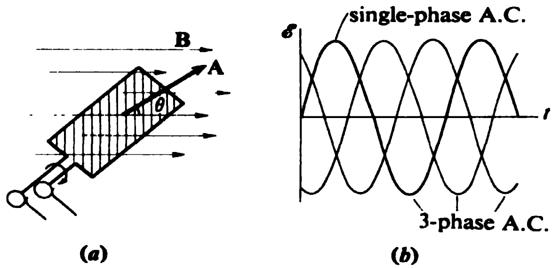


Fig. 9.6. (a) Simple A.C. generator; (b) the form of single-phase and 3-phase alternating current.

by windings on concave pole-pieces and the rotating coils are wound in slots on a steel drum armature which completes the magnetic circuit (section 14.7), the generator is of type III in table 9.1. In larger A.C. generators the field and armature windings are on the rotor and stator respectively for reasons mentioned above (type IV). Both types III and IV will behave as motors when supplied with A.C. but their speed of rotation is limited to that which would generate the frequency of the input if acting as a generator.

The reader is assumed to be familiar with the simple commutator which reverses every other half-cycle of the output of a single-coil A.C. generator of Fig. 9.6a. By winding sufficient coils on the armature and arranging commutator segments so that brushes take off current from several coils in series the output can be made approximately constant—D.C. with a slight ripple—and it can be used to excite its own field coils. The field coils are wound either in series with the armature windings, in parallel with them, or both. The armature current produces its own field which combines with that of the field windings to produce undesirable effects at the brush contacts. These can be eliminated by rotating the brushes slightly or, more commonly, by using small pole-pieces (interpoles) between the main ones and winding them in series with the armature: their field then compensates that of the armature windings. These are

type II generators: they will also form D.C. motors, or the small A.C. motors found on domestic machinery.

Induction Motors. These are the most common form of powerful A.C. motor. The principle is that, just as a coil rotating in a magnetic field has a current induced in it upon which a torque is exerted, so a rotating magnetic field should exert a torque on a stationary coil. The rotating magnetic field is produced by 3-phase A.C. in the stator coils and the rotor is an iron cylinder, completing the magnetic circuit, with copper bars embedded along the side connected at the ends by copper rings forming the 'squirrel cage'. The rotor never quite achieves the angular velocity of the magnetic field, for then the torque would vanish.

9.5 Self-Inductance

In the absence of magnetic materials, the flux density at any point due to a single circuit of given shape is proportional to the current in it, since the current is measured by its magnetic effect. It follows that the total flux Φ linking a circuit due to its own current I is proportional to I. We define the self-inductance L of a circuit as the ratio of Φ to I so that

$$\Phi = LI \qquad \text{(Definition of } L) \qquad (9.14)$$

and L is thus a constant for a circuit of given shape and size. The self-flux of a circuit defines the positive direction for Φ so that L is essentially positive. Its unit, the weber/amp, is called the *henry*, symbol H.

Self-induction. When the current in a rigid circuit changes, the induced e.m.f., by (9.9), is

$$\mathscr{E} = -L\,\mathrm{d}I/\mathrm{d}t \qquad (9.15)$$

the negative sign indicating that when I increases, $\mathscr{E}$ is negative and is thus in a direction opposing the growth of I (Fig. 9.7a). When a device is constructed to embody self-inductance it is known as an *inductor* and has the symbol in Fig. 9.7b in a circuit diagram, crossed by an arrow if variable and with parallel lines adjacent to it if it has a ferromagnetic core.

In a network with varying currents, an inductor acts as an e.m.f. $-L\,\mathrm{d}I/\mathrm{d}t$ with an internal resistance which cannot be avoided. Like a cell, therefore, it can be represented by an equivalent network consisting of a pure inductance L in series with its resistance R as in Fig. 9.7c. The pure inductance has no resistance and thus has a

potential difference across its ends equal to the e.m.f.: once the direction of I has been chosen arbitrarily in any problem, the potential drop across an inductor L is $L\,\mathrm{d}I/\mathrm{d}t$ just as that across R is RI (Fig. 9.7c).

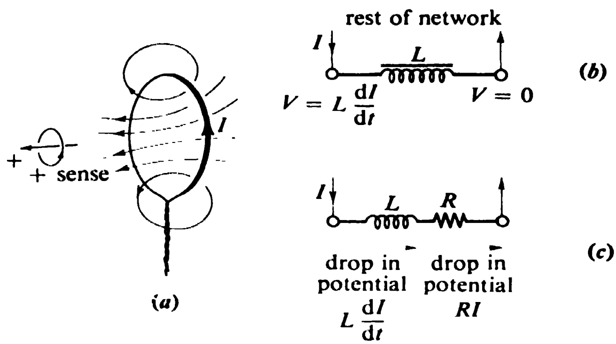


Fig. 9.7. (a) Self-flux of a circuit: increase of flux due to an increase of I causes e.m.f. in a negative sense and this opposes I; (b) potential drop across L; (c) equivalent circuit of self-inductor with resistance.

Growth and Decay of Current in an Inductor. In Fig. 9.8 the resistance R includes the internal resistance of the cell and of the inductor as well as any other resistance in the circuit. If I is the current flowing at any time t after the key K_1 is closed then we again

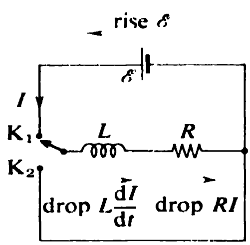


Fig. 9.8. Potential differences at any instant when current grows in a self-inductance with resistance.

assume that, in spite of varying currents, the steady current laws apply at any instant (cf. the treatment in section 6.5). Equating the total potential drop round the circuit to zero gives

$$\mathcal{E} - L\,\mathrm{d}I/\mathrm{d}t - RI = 0 \qquad (9.16)$$

or

$$L\,\mathrm{d}I/\mathrm{d}t + RI = \mathcal{E} \qquad (9.17)$$

which has the same form as (6.22) and is solved in the same way. The variable is now I and the initial condition is $I=0$ when $t=0$ because any instantaneous increase in I would mean that dI/dt was infinite and this would produce an infinite back e.m.f. in L. Thus

$$I = I_0(1-e^{-Rt/L}) \tag{9.18}$$

where $I_0=\mathscr{E}/R$ and is the current in the circuit when $t\to\infty$.

When the key is connected to K_2 the current is given by

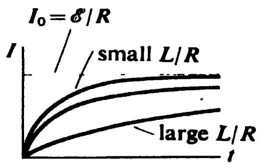
$$L\,dI/dt+RI = 0 \tag{9.19}$$

which, with the initial condition $I=I_0$ when $t=0$, gives

$$I = I_0e^{-Rt/L} \tag{9.20}$$

The growth (9.18) and the decay (9.20) of the current are plotted in Fig. 9.9. The time L/R is the time constant or relaxation time of the circuit just as CR was in section 6.5.

(a)



(b)

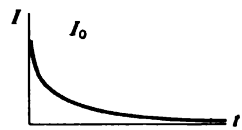


Fig. 9.9. (a) Growth and (b) decay of current in the network of Fig. 9.8.

Because *any* circuit has self-inductance, a broken contact when a current is flowing is equivalent to the introduction of a very high resistance. The time constant L/R thus suddenly decreases to a very small value and the *rate of collapse* of current is very high even if the current itself is small. The resultant induced e.m.f. is often sufficient to cause a spark to jump across the gap. A quantitative analysis is more complex than at first sight appears, because of the capacitance also introduced.

9.6 Calculation of Self-Inductance

In simple cases, L can be calculated from (9.14).

L of a Solenoid. If the turns of a helically wound solenoid are close enough and if the length is much greater than the diameter, the flux density to a good degree of accuracy is $\mu_0 nI$ over the whole cross-section from (8.7), n being the number of turns per unit length. The flux across each turn of area A is thus $\mu_0 nIA$ and,

because there are nl turns, the total self-flux is $\mu_0 n^2 IlA$. Hence

$$L_\infty = \mu_0 n^2 lA = \mu_0 N^2 A/l \tag{9.21}$$

N being the total number of turns. This gives the self-inductance of an ideal solenoid considered as part of an infinite solenoid. Finite solenoids have $L = bL_\infty$ where b is a function of the ratio of diameter to length which can only be obtained by more advanced methods. Table 9.2 gives values of b for various coils from which it is seen that (9.21) is excellent for yielding an order of magnitude.

Table 9.2

diameter/length	0	0·5	1·0	10·0
b	1	0·8	0·7	0·2

Distributed Inductance. Transmission lines were found in section 5.2 to possess distributed capacitance and it is clear from Fig. 9.10 that either type of line will also possess self-inductance. We first obtain a value for L by neglecting the flux within the wires or conductors themselves. For the twin cable of Fig. 9.10a, the flux

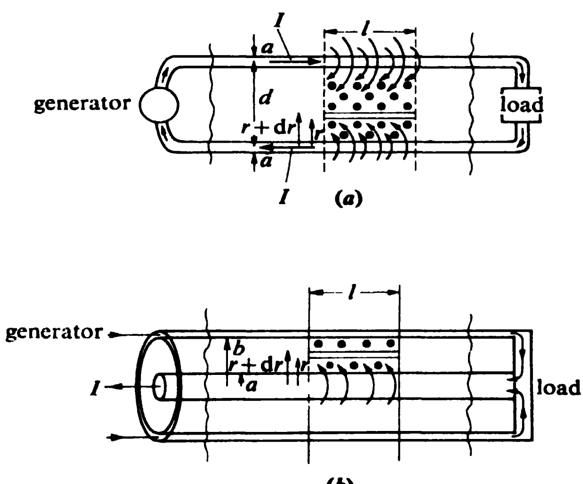


Fig. 9.10. Magnetic flux in (a) a twin cable, (b) a coaxial cable.

density at r from one wire is $\mu_0 I/2\pi r + \mu_0 I/2\pi(d-r)$. The flux across an elementary strip is the flux density multiplied by the area $l\,\mathrm{d}r$ and the total flux is

$$\Phi = \int_a^{d-a} \mu_0 I[\mathrm{d}r/r + \mathrm{d}r/(d-r)]/2\pi \quad \text{per unit length}$$

and hence

$$L = \frac{\mu_0}{\pi} \log_e \left(\frac{d-a}{a}\right) \quad \text{per unit length} \tag{9.22}$$

$$\doteqdot \frac{\mu_0}{\pi} \log_e \frac{d}{a} \quad \text{per unit length if } d \gg a$$

for a twin cable.

In a similar way

$$L = \frac{\mu_0}{2\pi} \log_e \frac{b}{a} \quad \text{per unit length} \tag{9.23}$$

for a coaxial cable (Fig. 9.10b), for here the current in the outer conductor produces no flux inside it.

Skin Effect. The neglect of flux within the conductors in the above calculations would be valid if all the current flowed in the outer surfaces of the twin wires and in the adjacent surfaces of the coaxial cable. When the current alternates at high frequency this condition in fact arises as follows and is known as the *skin effect*. Figure 9.11 shows a current-carrying loop of wire drawn with an exaggerated cross-section for clarity. If the current is steady, the elementary filament in the centre of the wire (mid-way between nos. 1 and 2) will link most flux, nos. 1 and 2 will link less, while nos. 3 and 4 will link only the flux external to the wire. If the current changes in time, the induced e.m.f. opposing the change is greatest in the centre filament and least in filaments 3 and 4. For alternating e.m.f.s, the effect is to create a greater impedance (section 10.4) to current flow along the centre than along the surface, with the result that more current tends to flow in the surface. The effect is very marked at high frequencies and makes (9.22) and (9.23) highly accurate.

Self-inductance of a Wire. At lower frequencies the current will be more evenly distributed and the flux in a wire will contribute to self-inductance. We cannot take the easy way out and assume that if the diameter of the wire is negligible the flux in the wire may be neglected, for the flux density B near the surface given by $\mu_0 I/2\pi r$

(8.11) tends to infinity as r tends to zero and the flux linked by the circuit, and hence the self-inductance, also become infinite (see also (9.22) and (9.23)).

However, assuming that the cross-section remains finite and the current density across it uniform, we can divide the current into infinitesimal filaments for each of which the self-flux tends to zero. We then calculate the flux linked by every filament due to the current in all the others and average the L's of the filaments so obtained over the cross-section.

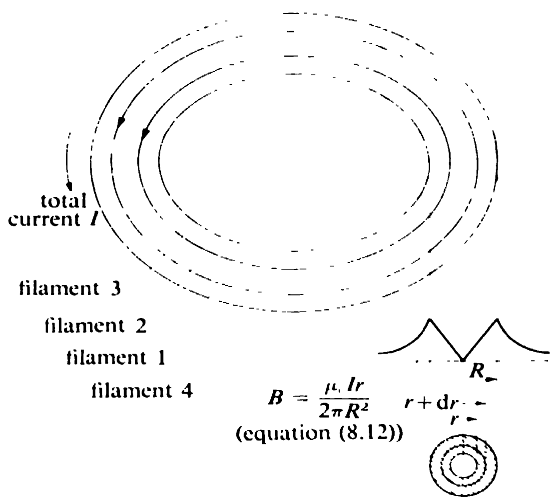


Fig. 9.11. A wire circuit carrying a current I.

In Fig. 9.11, all the filaments at a distance r from the central axis of the wire link the flux in the shaded area. By a method similar to that for the transmission lines, this is $\int_r^R \mu_0 Ixl\,dx/2\pi R^2$ or $\mu_0 Il(R^2 - r^2)/4\pi R^2$ for a length l of wire. The number of filaments linking this particular flux is a fraction $2\pi r\,dr/\pi R^2$ of the total filaments in the wire and hence the contribution of this bundle to the self-inductance is

$$\frac{\mu_0 l(R^2 - r^2)r\,dr}{2\pi R^4}$$

L is therefore the integral of this from 0 to R or

$$L = \mu_0 l/8\pi \tag{9.24}$$

Thus even a straight wire has a self-inductance of about 1/20 μH/m due solely to the flux in the wire. This will only apply when the skin effect is negligible and will even then often be small compared with the contribution from the flux outside. (We have neglected the permeability of the material of the wire, but this is only of importance when ferromagnetic—iron or nickel, for instance.)

Combinations of Self-inductors. It is easy to show from the definition that inductors in series have an equivalent inductance equal to the sum of the separate inductances provided there is no mutual flux linkage between one and another. Parallel and more complex combinations will be left until we deal with the general theory of circuits containing varying currents in chapter 10.

9.7 Mutual Inductance

When a current I_1 flows in one of two circuits as in Fig. 9.12 the flux Φ_2 linking the other, for a given geometry, is proportional to I_1 in the absence of magnetic materials so that

$$\Phi_2 = M_{12} I_1 \quad \text{(Definition of } M_{12}) \tag{9.25}$$

where M_{12} is a constant for the pair of circuits in their specified positions known as their *mutual inductance* and measured, like L,

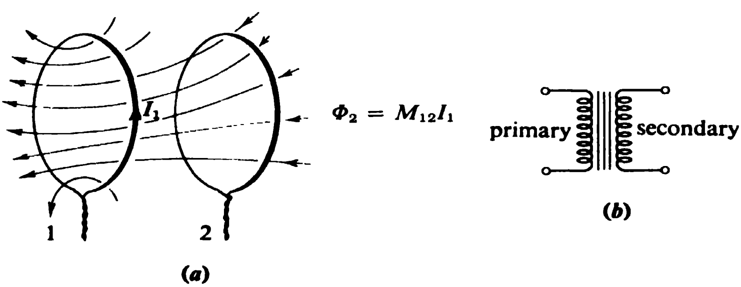


Fig. 9.12. (a) Flux linking circuit 2 due to I_1 in circuit 1; (b) a mutual inductance as a circuit element.

in henries. Similarly, a current I_2 in 2 produces a flux linkage

$$\Phi_1 = M_{21} I_2 \quad \text{(Definition of } M_{21}) \tag{9.26}$$

and we shall later show that $M_{12} = M_{21}$.

The sign given to the flux linking a circuit has been taken as positive when in the same direction as any self-flux. If there is no self-flux, then only the external flux can be used to allocate a sign, taken conventionally as positive. It follows that, if currents flow in both of two circuits, M may be positive or negative while if a steady current flows in only one circuit M is essentially positive.

Mutual Induction. If the current I_1 changes, the induced e.m.f. in circuit 2 is given by

$$\mathscr{E}_2 = -M_{12}\frac{\mathrm{d}I_1}{\mathrm{d}t} \tag{9.27}$$

the significance of the negative sign being indicated in Fig. 9.2b. As a circuit element, a *mutual inductor* is represented as in Fig. 9.12b, the parallel lines indicating an iron core if present. With such a core the inductor is usually a transformer (sections 10.7 and 10.8) one coil of which is used for the input of power (the primary) and the other for the output (the secondary).

In circumstances where the circuits are in relative motion and I_1 is simultaneously varying with time,

$$\mathscr{E}_2 = -M_{12}\frac{\mathrm{d}I_1}{\mathrm{d}t} - I_1\frac{\mathrm{d}M_{12}}{\mathrm{d}t} \tag{9.28}$$

a combination of changing-field and motional e.m.f.s.

A known current I reversed in the primary of a standard mutual inductor gives a standard flux change

$$\delta\Phi = 2MI \tag{9.29}$$

in the secondary, and this can be used to calibrate a fluxmeter or ballistic galvanometer, the charge flowing in the secondary circuit being

$$Q = 2MI/R \tag{9.30}$$

from (9.11).

9.8 Magnetic Energy 1: In Terms of L, M and I

A set of wire circuits carrying currents possesses more energy than the same set with no currents because of the work done in establishing them against induced e.m.f.s. We shall calculate this energy in stages, first in a single circuit, next in two circuits and finally in a set.

Magnetic Energy due to Self-inductance. Consider the growth of current I in a circuit containing self-inductance L and resistance R as in section 9.5. The rate at which energy is being transferred from the source of e.m.f. to the rest of the circuit at any instant is $I\mathscr{E}$ and from (9.17) we see that this is equal to $LI\,\mathrm{d}I/\mathrm{d}t + RI^2$. The second term is recognizable as the rate of consumption of energy in the resistance and it follows that the first represents the rate of storage in the inductance. In a time $\mathrm{d}t$, therefore, the energy stored is $LI\,\mathrm{d}I$ and, integrating from $I=0$ to $I=I_0$, the final current, gives

$$U_M = \tfrac{1}{2}LI_0^2 \tag{9.31}$$

The justification for calling the energy *stored* is that, when the current decays with no input of energy from $\mathscr{E}$, a further amount of energy $\frac{1}{2}LI_0^2$ is dissipated in R, as the reader can verify by calculating $\int_0^\infty RI^2\,\mathrm{d}t$ using (9.20).

Magnetic Energy due to Mutual Inductance. Consider now two circuits with self-inductance L_1 and L_2 and a mutual inductance M_{21} as defined in (9.26). First insert an e.m.f. in circuit 1 so that the current in it grows as above to its final value I_1, keeping any induced currents in 2 at zero by extra e.m.f.s which do no work since $I_2=0$. The energy stored, as above, is $\frac{1}{2}L_1I_1^2$.

Now insert an e.m.f. in circuit 2 to make the current in it grow to its final value I_2. When it has done so, naturally an energy $\frac{1}{2}L_2I_2^2$ is stored in 2, but while I_2 is changing an extra e.m.f. $-M_{21}\,\mathrm{d}i_2/\mathrm{d}t$ is induced in circuit 1, i_2 being the instantaneous value. We insert in 1 an adjustable e.m.f. of value $+M_{21}\,\mathrm{d}i_2/\mathrm{d}t$ thus keeping I_1 constant (and not upsetting the $\frac{1}{2}L_1I_1^2$ already stored). This time, however, the extra e.m.f. passes the current I_1 and thus puts energy into the system at the rate $I_1M_{21}\,\mathrm{d}i_2/\mathrm{d}t$. The total energy put in by this extra e.m.f. is therefore $I_1M_{21}\,\mathrm{d}i_2$ integrated from 0 to I_2, or $M_{21}I_1I_2$.

Hence for a pair of circuits

$$U_M = \tfrac{1}{2}L_1I_1^2 + M_{21}I_1I_2 + \tfrac{1}{2}L_2I_2^2$$

and because the same situation can be reached by allowing I_2 to grow first and finding that U_M is then $\frac{1}{2}L_1I_1^2 + M_{12}I_2I_1 + \frac{1}{2}L_2I_2^2$, it follows that $M_{12} = M_{21}$: there is only one mutual inductance M between two circuits, and the magnetic energy is

$$U_M = \tfrac{1}{2}L_1I_1^2 + MI_1I_2 + \tfrac{1}{2}L_2I_2^2 \tag{9.32}$$

This can also be written as

$$U_M = \tfrac{1}{2}I_1(L_1I_1 + MI_2) + \tfrac{1}{2}I_2(L_2I_2 + MI_1) \tag{9.33}$$

$$= \tfrac{1}{2}I_1\Phi_1 + \tfrac{1}{2}I_2\Phi_2 \tag{9.34}$$

where Φ_1 and Φ_2 are *total* magnetic fluxes linking each circuit.

Magnetic Energy in a Set of Circuits. For any number of circuits with self-inductances $L_1, L_2, \ldots, L_i, \ldots$, and mutual inductances $M_{12}, M_{13}, M_{23}, \ldots, M_{ij}, \ldots$, equation (9.32) can be extended to

$$U_M = \sum_i \tfrac{1}{2}L_iI_i^2 + \sum_i \sum_j M_{ij}I_iI_j \tag{9.35}$$

and by splitting each M term into two halves and grouping as in (9.33)

$$U_M = \sum_i \tfrac{1}{2}I_i\Phi_i \tag{9.36}$$

It is easy to show that for small increments of all the currents by dI_i, the increment in U_M is

$$dU_M = \Sigma\, \tfrac{1}{2}I_i\, d\Phi_i = \Sigma\, \tfrac{1}{2}\Phi_i\, dI_i \tag{9.37}$$

because I and Φ are linearly related (but see section 14.5).

9.9 Magnetic Energy 2: In Terms of B

We now express U_M in terms of B in a way similar to that in which we expressed U_E in terms of E in section 5.6. To do this, consider a single circuit carrying a current I linked by flux (Fig. 9.13). The

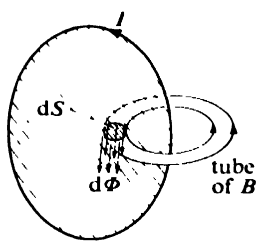


Fig. 9.13. Calculation of the magnetic energy of a current-carrying circuit in terms of B.

lines of B crossing a small area dS normally will form a closed tube as shown if continued and, because the number of lines does not change, the flux in the tube will remain the same however B and dS may change, such that $d\Phi = B\, dS$. If the total energy $U_M = \tfrac{1}{2}I\Phi$,

where Φ is the total flux linking the circuit, an amount $\frac{1}{2}I\,\mathrm{d}\Phi$ could be associated with the tube. By the circuital theorem (8.10), $I=\oint B\,\mathrm{d}s/\mu_0$ round any closed path so that

$$\mathrm{d}U_M = \tfrac{1}{2}\,\mathrm{d}\Phi \oint B\,\mathrm{d}s/\mu_0 = \tfrac{1}{2}\oint B^2\,\mathrm{d}\tau/\mu_0$$

is the energy associated with the tube, where $\mathrm{d}\tau$ is an element of volume. If we integrate over the two variables represented by $\mathrm{d}S$ so as to include all the tubes, we finally obtain

$$U_M = \iiint \frac{B^2}{2\mu_0}\,\mathrm{d}\tau \tag{9.38}$$

which could be expressed by saying that the energy density associated with a magnetic field is given by $B^2/2\mu_0$, although the same qualifications apply here as in the electric case concerning location of energy.

As an example, consider a portion of the long solenoid carrying a current I. With the usual notation, B is $\mu_0 nI$ everywhere inside and the volume is lA so that (9.38) gives $U_M=\frac{1}{2}\mu_0 n^2 lAI^2$: exactly equal to $\frac{1}{2}LI^2$.

9.10 Calculation of Mutual Inductance

In simple cases, mutual inductance may be calculated by finding the flux linked by *either* circuit due to unit current in the other: more powerful methods are developed in more advanced texts.

Coaxial Solenoids. A convenient form of fixed mutual inductance is a pair of coaxial solenoids as in Fig. 9.14. Using the

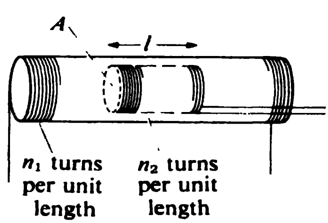


Fig. 9.14. A mutual inductor.

notation indicated therein, a current I in the longer solenoid produces a flux density $\mu_0 n_1 I$ which will be approximately constant over the whole length of the short solenoid, each turn of which links a flux $\mu_0 n_1 IA$. Hence

$$M = \mu_0 n_1 n_2 lA \tag{9.39}$$

Small Circular Coil on the Axis of a Large Coil. In Fig. 9.15, the small coil is situated in a flux density $B=\mu_0 N_1 I a_1^2/2(a_1^2+x^2)^{3/2}$

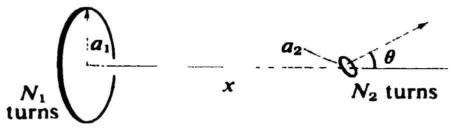


Fig. 9.15. ***A small coil on the axis of a larger one.***

when I flows in the large one. The flux linked is $N_2 \pi a_2^2 B \cos\theta$ and so

$$M = (\mu_0 N_1 N_2 \pi a_1^2 a_2^2 \cos\theta)/2(a_1^2+x^2)^{3/2} \tag{9.40}$$

Two Coupled Coils. The coils of Fig. 9.16 have a mutual inductance of magnitude M. If Q and R are connected and a current

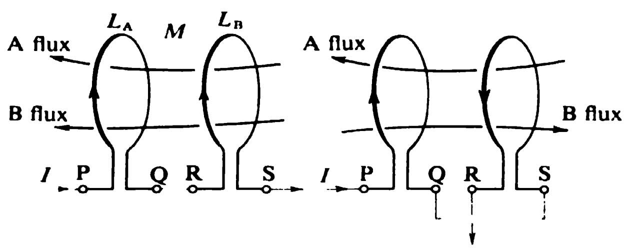


Fig. 9.16. ***Two coupled coils.***

I passed through both coils in series then, considered as one circuit, the total flux linkage is $L_A I + L_B I + 2MI$ and the self-inductance is

$$L_1 = L_A + L_B + 2M \tag{9.41}$$

If, on the other hand, Q and S are connected and a current passed from P to R the total flux linkage is now $L_A I + L_B I - 2MI$ (the self-fluxes must be positive and the mutual fluxes oppose them) so that

$$L_2 = L_A + L_B - 2M \tag{9.42}$$

The same effect could be obtained by retaining the connections but rotating one coil through 180°—a method which can therefore be used to obtain a self-inductance varying smoothly from L_1 to L_2. Alternatively, because $M=\frac{1}{4}(L_1-L_2)$ from (9.41) and (9.42), a

mutual inductance may be obtained from measurements only of self-inductance, which are often simpler.

A transformation we shall find useful in chapter 10 is illustrated in Fig. 9.17. Two coupled coils with self-inductances L_A and L_B and a mutual inductance M connected between three points A, B

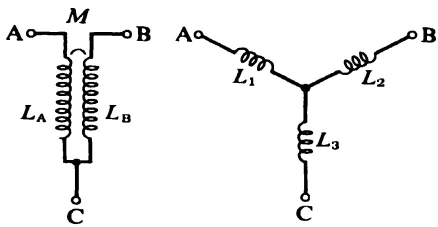


Fig. 9.17. These are equivalent if $L_1 = L_A \pm M$, $L_2 = L_B \pm M$, $L_3 = \mp M$.

and C are equivalent to the star or Y of L's if the effective inductances between the three points taken in pairs are equal.

$$\begin{aligned} \text{AB:} \quad & L_A + L_B \pm 2M = L_1 + L_2; \\ \text{BC:} \quad & L_B = L_2 + L_3; \\ \text{CA:} \quad & L_A = L_3 + L_1 \end{aligned}$$

Solved for L_1, L_2 and L_3, these yield

$$L_1 = L_A \pm M; \quad L_2 = L_B \pm M; \quad L_3 = \mp M \tag{9.43}$$

Coefficient of Coupling. Because the maximum flux from one circuit A which can link another, B, is the whole of its self-flux, M_{AB} cannot be greater than L_A; nor can it be greater than L_B. Hence

$$M_{AB} = k_1 L_A = k_2 L_B$$

where both k_1 and k_2 are $\leqslant 1$, and therefore

$$M_{AB}/\sqrt{L_A L_B} = k \tag{9.44}$$

where k is a constant for a given pair of circuits in given positions known as the *coefficient of coupling*, with a value which cannot exceed 1. Coils with k above about 0·5 are said to be tightly coupled.

9.11 Forces, Couples and Changes in Energy

We use a method similar to that in section 5.7 to find the force or couples between two current-carrying circuits, the difference being that in the electric case the sources of the field (the charges) remained constant without extra energy being needed. Here, the sources of field are currents and if these are to be constant during any changes, extra working e.m.f.s are needed.

Changes at Constant Current. In Fig. 9.18, the internal forces of attraction are shown dotted and the external forces, keeping the circuits apart and in equilibrium, in full. Circuit B is subject to an internal force **F**, which we are trying to find, and an external force

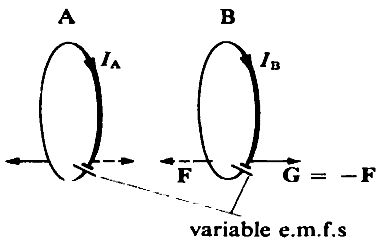


Fig. 9.18. Forces between current-carrying circuits. External forces maintaining equilibrium are shown full, internal forces dotted.

G = −**F**. For any displacement ds of B let extra e.m.f.s be introduced into the two circuits to keep I_A and I_B constant, and let the change in mutual inductance be dM. The increment in magnetic energy is

$$\mathrm{d}U_M = I_A I_B\,\mathrm{d}M \tag{9.45}$$

from (9.32). The extra e.m.f.s will be of magnitude $I_B\,\mathrm{d}M/\mathrm{d}t$ in A and $I_A\,\mathrm{d}M/\mathrm{d}t$ in B so that both will be supplying energy to the whole system at a rate $I_A I_B\,\mathrm{d}M/\mathrm{d}t$. Thus both will lose an amount of energy in time dt equal to $I_A I_B\,\mathrm{d}M$ so that the increment in the energy of the system (circuits + batteries) is $-I_A I_B\,\mathrm{d}M$ or $-\mathrm{d}U_M$. The work done *on* the circuit B by the external force is $\mathbf{G}\cdot\mathrm{d}\mathbf{s}$ or $-\mathbf{F}\cdot\mathrm{d}\mathbf{s}$ and therefore

$$-\mathbf{F}\cdot\mathrm{d}\mathbf{s} = -I_A I_B\,\mathrm{d}M = -\mathrm{d}U_M$$

or

$$F_s = I_A I_B \frac{\partial M}{\partial s} = \left(\frac{\partial U_M}{\partial s}\right)_I \tag{9.46}$$

Similarly the couple about any axis θ is given by

$$T_\theta = I_A I_B \frac{\partial M}{\partial \theta} = \left(\frac{\partial U_M}{\partial \theta}\right)_I \tag{9.47}$$

Changes at constant current thus mean that, of the energy supplied by the batteries, half goes to increasing the internal magnetic energy and half to doing external work. These changes correspond to those at constant potential in electrostatics. Changes at constant flux linkage (corresponding to constant charge) yield formulae like $F_s = -(\partial U_M/\partial s)_\Phi$ but they are in practice of little use, and give the same forces and couples in a static situation as do (9.46) and (9.47), as they must.

Potential Energy. In section 8.4 we saw that the potential energy of a current I linked by external flux Φ was $U = -I\Phi$ and that forces and couples could be obtained from this by $F_s = -(\partial U/\partial s)_I$, etc. The above example of two circuits is a special case of this for which $\Phi_B = MI_A$ and hence $U = -I_A I_B M$, and it is easily seen that (8.20) will give the same as (9.46) and (9.47) above.

Confusion often arises between the potential energy of two circuits

$$U = -MI_A I_B \tag{9.48}$$

from (8.19) and the magnetic energy U_M which includes a term $+MI_A I_B$. The same question is not asked in electrostatics because U_E and U are equal: why the difference? The answer lies in that electric energy depends solely on the *positions* of the charges and is thus identical with potential energy, whereas the magnetic energy of currents depends on both the positions and motion of charges and cannot be completely characterized as either potential or kinetic. When we calculate potential energy we find the work done by external mechanical forces in moving the circuits from infinity to their final positions *with constant currents*: this is bound to be different from the total energy needed to establish the currents in their final positions because, as we have seen, batteries must put in $+2MI_A I_B$ to keep the currents constant so that the total energy increases by $2MI_A I_B + U$ or $+MI_A I_B$, thus accounting for the + sign in U_M.

Example. The small coil of Fig. 9.15 will experience a force F_x

and a couple T_θ given by

$$F_x = \frac{3\mu_0 I_1 I_2 N_1 N_2 \pi a_1{}^2 a_2{}^2 x \cos\theta}{2(a_1{}^2+x^2)^{5/2}};$$
$$T_\theta = \frac{\mu_0 I_1 I_2 N_1 N_2 \pi a_1{}^2 a_2{}^2 \sin\theta}{2(a_1{}^2+x^2)^{3/2}} \tag{9.49}$$

using (9.40), (9.46) and (9.47).

9.12 Inductors in Practice

As we shall see in chapter 14, the effect of a medium filling the whole of the space round a coil is to change its inductance by a factor μ_r known as the *relative permeability* of the medium so that, if L_0 is the inductance *in vacuo* and L_m in the medium, $L_m = \mu_r L_0$. For nearly all materials μ_r is so close to 1 that the change in inductance is negligible, but for a limited class, described as ferromagnetic, μ_r is very large and is not constant as the magnetic field changes: thus although ferromagnetic materials allow us to obtain large inductances in a small volume, the linearity of L and M is sacrificed.

Standard inductors, whether fixed or variable, are air-cored and wound on formers in such a way that the positions of the windings are definite; temperature variations of inductance due to expansion are almost eliminated by compensating the increase in diameter by an increase in length (see equation (9.21) in which A and l can be made to increase in proportion by a suitable choice of materials). The N.P.L. primary standard of mutual inductance is described in section 16.2. Variable self-inductance decades can be constructed if mutual flux linkage is avoided, and variable mutual inductances are obtained by relative motion of a pair of coils.

Inductors for use in electronic power circuits will be iron-cored at low frequencies and air-cored at high frequencies because the reactance ωL (section 10.4) rather than L itself is the quantity required.

9.13 Summary of Chapter 9

Induced e.m.f.s are of two kinds: (1) changing-field e.m.f.s in stationary rigid circuits given by

$$\mathscr{E} = -\frac{\partial \Phi}{\partial t} \tag{9.9}$$

where Φ is the magnetic flux linked; and (2) motional e.m.f.s given by *either* the current element formula

$$d\mathscr{E} = (\mathbf{v}\times\mathbf{B})\cdot d\mathbf{l} = v_\perp B\, dl_\perp \quad \text{in magnitude} \tag{9.4}$$

or the flux-cutting law

$$\mathscr{E} = d\Phi/dt \qquad (\Phi \text{ is flux cut}) \tag{9.7}$$

or by the flux-linking rule

$$\mathscr{E} = -d\Phi/dt \qquad (\Phi \text{ is flux linked}) \tag{9.8}$$

whichever is convenient.

The important part of all this for the future is contained in (9.9) which can be written out in the form

$$\oint_C \mathbf{E}\cdot d\mathbf{l} = -\frac{\partial}{\partial t}\iint_S \mathbf{B}\cdot d\mathbf{S} = \iint_S -\frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{S} \tag{9.50}$$

S being the surface bounded by the closed path C. This means that the changing $\mathbf{B}$ through the surface S is accompanied by an electric field acting round the path C: the lines of $\mathbf{E}$ will thus, unlike those of an electrostatic field, close on themselves. We shall pursue this further in sections 11.4 and 15.1.

Returning to more detail, we have defined two important properties of networks, self-inductance and mutual inductance, in terms of flux linkage per unit current:

$$\Phi = LI; \qquad \Phi_2 = M_{12}I_1 \qquad \text{(9.14) and (9.25)}$$

and have used these definitions to calculate L and M in simple cases: all expressions have the form $\mu_0 \times$ a factor with the dimensions of length.

Magnetic energy in the absence of magnetic materials can be expressed as

$$U_M = \sum_i \tfrac{1}{2}L_iI_i^2 + \sum_i\sum_j M_{ij}I_iI_j = \sum_i \tfrac{1}{2}I_i\Phi_i = \iiint \frac{B^2}{2\mu_0}\, d\tau$$

(9.35), (9.36) and (9.38)

while forces and couples between two circuits are given by

$$F_s = I_A I_B \frac{\partial M}{\partial s}; \qquad T_\theta = I_A I_B \frac{\partial M}{\partial \theta} \qquad \text{(9.46) and (9.47)}$$

References

Faraday's work is very well documented, largely as a result of his own notebooks. Good introductions are to be found in Martin (1949) and Williams (1963), but his own *Experimental Researches in Electricity* (1839) is a masterly account of his methods.

PROBLEMS

SECTION 9.1

9.1 Show how Lenz's law can be used to find the direction of induced current flow in (*a*) a circular coil rotating about a diameter in a uniform magnetic field, (*b*) a stationary circular coil towards which a coaxial current-carrying coil is moving.

SECTION 9.2

9.2 Estimate the magnitude of the potential differences produced by motion of a conductor in the earth's magnetic field for which B is $3{\cdot}2 \times 10^{-5}$ Wb/m^2 and the angle of dip is $64°\ 9'$, taking as an example a car bumper of length 1 m travelling at 100 km/hour.

SECTION 9.3

9.3 A metal hoop of radius a and resistance R is held with its plane in the magnetic meridian, and falls over through 90° to the east. Find the total quantity of charge which flows while the hoop is falling if the earth's field has a flux density B and an angle of dip δ. Evaluate the charge for $a = \frac{1}{2}$m, $R = 0{\cdot}2\ \Omega$ and the earth's field as in problem 9.2.

9.4 A brass hoop of radius a and resistance R is placed with its plane perpendicular to a uniform magnetic field whose flux density fluctuates. If $B = B_0 \sin \omega t$, what is the induced current in the hoop?

*9.5 A brass disc of radius a, thickness b and conductivity σ has its plane perpendicular to a uniform magnetic field whose flux density varies according to $B = B_0 \sin \omega t$. Assuming that the eddy currents flow in concentric circles about the centre of the disc, find the total current flowing at any instant and the mean power dissipated as heat. Comment on the result as an indication of the factors affecting eddy current losses in iron.

SECTION 9.4

9.6 The Faraday disc of Fig. 9.1 is to be used as a motor by including a battery in the circuit shown. If the current flowing is I and if the magnetic flux density **B** is uniform, show that the torque exerted on the disc is $\Phi I/2\pi$ where Φ is the flux crossing the whole disc. Find I if the battery e.m.f. is $\mathscr{E}$ and the resistance of the circuit is R, and account for the power consumed from the battery.

9.7 Deduce equation (9.12) by considering the forces on the electrons in the rotating disc.

9.8 In the simple A.C. generator in which the e.m.f. is given by (9.13), show that the torque opposing the motion is $\Phi_m I \sin \omega t$ where I is the armature current.

9.9 If the A.C. generator of problem 9.8 is converted to a D.C. motor by including a battery of e.m.f. $\mathscr{E}$ in series with the coil and using a simple commutator, find the current in the coil when rotating at an angular velocity ω and hence find the mechanical power output. Assume that the flux density is constant.

9.10 A shunt-wound motor operates on a 100 V D.C. supply and takes a current of 5 A. If the field and armature windings have resistances of 200 Ω and 1 Ω respectively, find the back e.m.f. developed in the armature windings and the efficiency of the motor.

SECTION 9.5

9.11 In the circuit of Fig. 9.8 find expressions for the potential differences across L and R subsequent to the closing of K_1 and plot the variation against time.

SECTION 9.6

9.12 Use the fact that all expressions for self-inductance have the form $\mu_0 \times$ *length* to find the relation between 1 H and 1 e.m.u. of inductance.

9.13 Estimate the order of magnitude of the self-inductance of an air-cored solenoid of length 20 cm with one layer of 10 turns per cm, each turn forming a circle of radius 2 cm.

SECTION 9.7

9.14 The self-inductance of a coil is 20 mH. When a current flows in it, $\frac{1}{5}$ of its magnetic flux links a second coil. What e.m.f. is induced in the second coil if a current in the first collapses at a uniform rate of 0·5 A s?

SECTION 9.8

9.15 A stationary circuit carries a current I_1 while a second circuit carrying I_2 moves towards it, the mutual inductance at any instant being M. What power must be provided to maintain I_1 and I_2 constant if the two circuits are rigid?

SECTION 9.10

9.16 Two coils A and B have self-inductances L_1 and L_2. When a steady current flows in A, a quarter of the magnetic flux links B. Find the proportion of the flux from a current in B which links A, and evaluate the coefficient of coupling.

9.17 Find the mutual inductance between an infinitely long straight wire and a one-turn rectangular coil whose plane passes through the wire and

two of whose sides are parallel to the wire. The sides parallel to the wire are of length a, the other sides are of length b and the side nearest the wire is a distance d from it.

SECTION 9.11

9.18 Find the force on the coil in problem 9.17 when a current I flows in both circuits. Check the result by using (8.24).

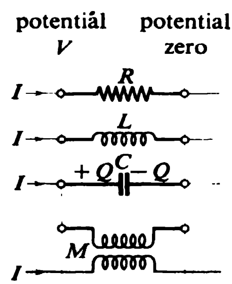
CHAPTER 10

VARYING CURRENTS IN LINEAR NETWORKS

Any network of conductors is bound to possess capacitance, resistance and inductance either concentrated in circuit elements or distributed as in connecting leads, cables and transmission lines. The properties of general networks containing either steady sources of e.m.f. producing transient changes (sections 10.2 and 10.3) or alternating sources (sections 10.4 onwards) are important practically but can be derived entirely from principles already established, together with the assumption of quasi-steady conditions.

10.1 Circuit Elements

The relations between potential difference and current for pure or ideal circuit elements are shown in Fig. 10.1 in which the assumed



$V = RI$

$V = L\dfrac{\mathrm{d}I}{\mathrm{d}t}$

$V = \dfrac{Q}{C}$; $I = \dfrac{\mathrm{d}Q}{\mathrm{d}t}$

$V = M\dfrac{\mathrm{d}I}{\mathrm{d}t}$

Fig. 10.1. Circuit elements.

directions for current should be carefully noted. If the current has to be chosen so that it flows away from the assumed positive plate of the condenser then $I = -\mathrm{d}Q/\mathrm{d}t$. There may also be reasons why M should be assumed negative (section 9.7).

The following assumptions will initially be made in solving network problems:

1. The steady current laws for conducting circuits from chapter 6 (summarized by $\Sigma I = 0$ at a junction and $\Sigma V = 0$ round a mesh) apply at any instant.
2. The circuit elements are all linear: that is, R, C, L and M are independent of the magnitude of current or charge. Many components are in practice non-linear—resistors because of the temperature variation, inductors because of ferromagnetic cores, condensers because of non-linear dielectrics—but for small *variations* of current the non-linearity can usually be neglected. We shall not be concerned here with grossly non-linear elements (but see problems 10.1 and 10.2).
3. All sources of e.m.f. have no internal resistance, capacitance or inductance unless explicitly quoted, and the amplitude of the e.m.f. generated is independent of the load placed on it.

The results of these three assumptions and the relations of Fig. 10.1 is to yield for any network sets of differential equations which are linear and have constant coefficients such as

$$A_n \frac{\mathrm{d}^n y}{\mathrm{d}t^n} + A_{n-1} \frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}} + \cdots + A_1 \frac{\mathrm{d}y}{\mathrm{d}t} + A_0 y = \mathrm{f}(t) \qquad (10.1)$$

where the function $\mathrm{f}(t)$ represents the applied e.m.f.s. We shall first consider circuits containing only steady sources or no sources, which makes $\mathrm{f}(t)$ effectively zero (section 10.2) and equation (10.1) a *homogeneous* equation. Then we shall look at circuits containing sinusoidally varying e.m.f.s, making $\mathrm{f}(t) = \mathscr{E}_0 \sin \omega t$ and equation (10.1) a *non-homogeneous* equation. Finally we deal with the effects of applying periodic but non-sinusoidal e.m.f.s.

In all cases the general solution of an equation such as (10.1) is any solution containing n constants, which are determined by the initial conditions (e.g. the first order equation (6.22) had a solution with one constant determined by one initial condition).

10.2 Transients in a Series *LCR* Circuit

The growth and decay of charge and current in CR and LR circuits (sections 6.5 and 9.5) are examples of *transient* effects in that all changes tend to disappear as $t \to \infty$. We now consider the circuits of Fig. 10.2 which represent the situations at a time t after closing a key to begin the charge or discharge of C through R and L.

Application of $\Sigma V = 0$ and $I = \mathrm{d}Q/\mathrm{d}t$ to the charging of C in Fig. 10.2a yields

$$L\frac{\mathrm{d}^2Q}{\mathrm{d}t^2} + R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{Q}{C} = \mathscr{E} \tag{10.2}$$

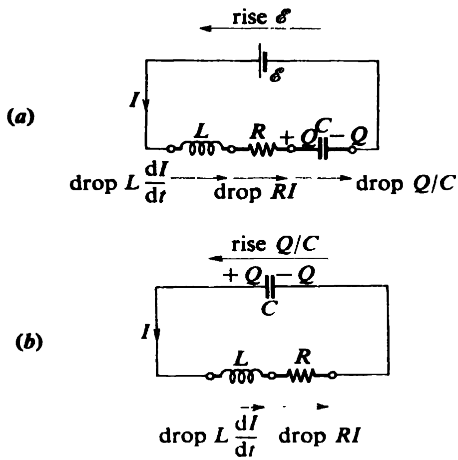


Fig. 10.2. (a) Charge and (b) discharge of a condenser through a resistance and inductance.

Let $\mathscr{E}C = Q_0$, a constant to be interpreted later, when

$$L\frac{\mathrm{d}^2(Q-Q_0)}{\mathrm{d}t^2} + R\frac{\mathrm{d}(Q-Q_0)}{\mathrm{d}t} + \frac{(Q-Q_0)}{C} = 0 \tag{10.3}$$

a differential equation of the form of (10.1) with $n = 2$, $\mathrm{f}(t) = 0$ and the variable $y = (Q - Q_0)$. The form of the solution of (6.22) suggests that we try as a solution

$$Q - Q_0 = A\mathrm{e}^{\lambda t}$$

and this substituted in (10.3) gives

$$L\lambda^2 + R\lambda + 1/C = 0 \tag{10.4}$$

Hence λ has two values:

$$\lambda_1 = -\alpha + (\alpha^2 - \omega_0^2)^{1/2}, \qquad \lambda_2 = -\alpha - (\alpha^2 - \omega_0^2)^{1/2}$$

where we define

$$\alpha = R/2L; \qquad \omega_0 = 1/(LC)^{1/2} \tag{10.5}$$

The solution of (10.3) is thus

$$Q-Q_0 = A'e^{\lambda_1 t}+B'e^{\lambda_2 t}; \qquad I = dQ/dt \tag{10.6}$$

where A' and B' are constants determined by the initial conditions $I=0$, $Q=0$ at $t=0$ (see section 9.5 for the reason why $I=0$). Equation (10.6) is a *general* solution of (10.3) because (*a*) it is a solution and (*b*) it contains two constants of integration provided $\lambda_1 \neq \lambda_2$ (for $\lambda_1=\lambda_2$ see 'critical damping' below).

Turning now to the discharge, where the original charge on C is Q_0, the application of $\Sigma V=0$ and $I=-dQ/dt$ yields

$$L\frac{d^2Q}{dt^2}+R\frac{dQ}{dt}+\frac{Q}{C} = 0 \tag{10.7}$$

which is of the same form as (10.3) except that the variable is Q. The general solution to this is

$$Q = Ae^{\lambda_1 t}+Be^{\lambda_2 t}; \qquad I = -dQ/dt \tag{10.8}$$

where λ_1 and λ_2 are as above but A and B are determined by the initial conditions $I=0$, $Q=Q_0$ when $t=0$.

We shall carry through the calculation for the discharge only: the solution for the charging of C will be very similar.

Overdamped Circuit, $CR^2/4L>1$. Here $\alpha^2-\omega_0^2$ is positive, λ_1 and λ_2 are real and the solutions are exponential. The initial conditions substituted in (10.8) give

$$A+B = Q_0; \qquad \lambda_1 A+\lambda_2 B = 0$$

from which A and B can be found. The charge and current are thus

$$Q = \frac{Q_0}{\lambda_1-\lambda_2}(-\lambda_2 e^{\lambda_1 t}+\lambda_1 e^{\lambda_2 t}) \tag{10.9}$$

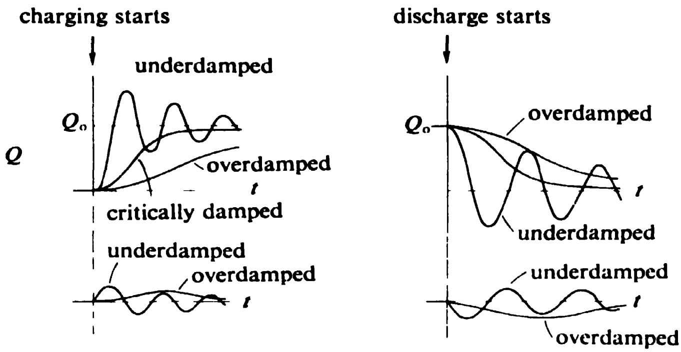
$$I = \frac{Q_0\lambda_1\lambda_2}{\lambda_1-\lambda_2}(-e^{\lambda_1 t}+e^{\lambda_2 t}) \tag{10.10}$$

As $t\to\infty$, the fact that λ_1 and λ_2 are both negative means that I tends to zero and is therefore transient. Substitution of the values for λ_1 and λ_2 into (10.10) gives

$$I = \frac{Q_0\omega_0^2 e^{-\alpha t}}{2(\alpha^2-\omega_0^2)^{1/2}}\left(e^{(\alpha^2-\omega_0^2)^{1/2}t}-e^{-(\alpha^2-\omega_0^2)^{1/2}t}\right)$$

$$= Q_0\omega_0^2 e^{-\alpha t}\frac{\sinh\omega' t}{\omega'} \tag{10.11}$$

where $\omega' = (\alpha^2 - \omega_0{}^2)^{1/2}$. The same variation of current is obtained on charging C and the curves of I and Q against t are shown in Fig. 10.3.



(*a*) (*b*)

***Fig. 10.3.** Variation of Q and I with time for (a) charge and (b) discharge of C through L and R.*

Underdamped Circuit, $CR^2/4L < 1$. Here $\alpha^2 - \omega_0{}^2$ is negative, λ_1 and λ_2 are complex and, although the solutions (10.9) to (10.11) apply, its form is clearer if we put $(\alpha^2 - \omega_0{}^2)^{1/2} = j\omega_N$ where $j^2 = -1$, so that

$$\omega_N{}^2 = \omega_0{}^2 - \alpha^2 \tag{10.12}$$

and

$$I = Q_0\omega_0{}^2 e^{-\alpha t} \frac{\sin \omega_N t}{\omega_N} \tag{10.13}$$

using either $e^{j\omega t} = \cos \omega t + j \sin \omega t$ or $\sinh j\omega t = j \sin \omega t$. The variation of Q in discharge is

$$Q = Q_0\omega_0 e^{-\alpha t} \frac{\sin(\omega_N t + \delta)}{\omega_N} \tag{10.14}$$

where $\tan \delta = \omega_N/\alpha$. Equations (10.13) and (10.14) are those of damped harmonic motion and are plotted in Fig. 10.3.

Critically Damped Circuit, $CR^2/4L = 1$. Here the roots of (10.4) are both $-\alpha$ and (10.8) becomes $(A+B)e^{-\alpha t}$ which contains only one constant and cannot be a general solution. In this case we try

$$Q = (A + Bt)e^{-\alpha t}$$

which satisfies (10.7) when $CR^2/4L=1$. The initial conditions give $A=Q_0$ and $B=\alpha A$ for the discharge and therefore

$$Q = Q_0\left(1+\frac{Rt}{2L}\right)e^{-Rt/2L} \tag{10.15}$$

and

$$I = -\mathrm{d}Q/\mathrm{d}t = -Q_0\frac{Rt}{2L}e^{-Rt/2L} \tag{10.16}$$

which are also plotted in Fig. 10.3.

Analogy with Mechanical Systems. The motion of a particle of mass m attracted to an origin by a force μx and subjected to a resistance $k\,\mathrm{d}x/\mathrm{d}t$ is governed by the differential equation

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}+k\frac{\mathrm{d}x}{\mathrm{d}t}+\mu x = 0 \tag{10.17}$$

which is of exactly the same form as (10.7) and has similar solutions. The analogy is expressed by the equivalents

$$x \leftrightarrow Q,\quad \frac{\mathrm{d}x}{\mathrm{d}t} \leftrightarrow I,\quad m \leftrightarrow L,\quad k \leftrightarrow R,\quad \mu \leftrightarrow \frac{1}{C} \tag{10.18}$$

and goes very deep. Just as mass is a measure of the resistance to changes in motion of the particle, so L is a measure of the inertia of the circuit opposing changes in current. Without the middle terms, the oscillations are harmonic with energy conserved but continuously changing from kinetic, $\frac{1}{2}mv^2$ or $\frac{1}{2}LI^2$, to potential, $\frac{1}{2}\mu x^2$ or $\frac{1}{2}Q^2/C$, and back again. The middle terms contain the factors responsible for the dissipation of energy from the system occurring at a rate $k(\mathrm{d}x/\mathrm{d}t)^2$ or RI^2, while damped harmonic oscillations take place if $k^2<4\mu/m$ or $R^2<4L/C$.

Natural Frequency and Logarithmic Decrement. The natural angular frequency of damped oscillations is, from (10.12),

$$\omega_N = (\omega_0{}^2-\alpha^2)^{1/2} = \left(\frac{1}{LC}-\frac{R^2}{4L^2}\right)^{1/2} \tag{10.19}$$

The *decrement* of decaying oscillations, defined* as the ratio of successive maxima, is

$$\Delta = e^{-\alpha t}/e^{-\alpha(t+T)} = e^{\alpha T}$$

* Some prefer to define Δ as the ratio of a maximum to the next minimum so that their Δ is only half of the one we use here.

where T is the period $2\pi/\omega_N$. The *logarithmic decrement*, Λ, is $\log_e \Delta$ and so

$$\Lambda = \alpha T = \pi \frac{R}{\omega_N L} \tag{10.20}$$

$$= 2\pi/(4L/CR^2-1)^{1/2} \tag{10.21}$$

10.3 Transients in Coupled Circuits

Networks more complex than that of the last section can be solved in a similar way, but when more than one mesh occurs the currents are governed by a set of simultaneous differential equations whose solutions are usually quite complicated. An important class of network is one with two meshes each with its own natural frequency and possessing a common element which couples them together. Our main interest lies in oscillations with small or negligible damping and, because the presence of resistance complicates formulae without greatly affecting the main results, we shall ignore it.

A common and important coupling element is the mutual inductance. Consider the network of Fig. 10.4 in which C_1 with an

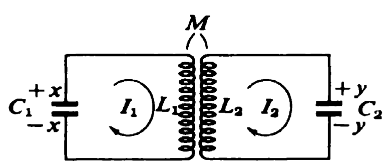


Fig. 10.4. Mutually coupled circuits.

initial charge Q_0 is discharging, the situation illustrated being that at time t. The mesh currents are $I_1 = -\mathrm{d}x/\mathrm{d}t$, $I_2 = +\mathrm{d}y/\mathrm{d}t$, and by applying $\Sigma\ V = 0$ to each mesh:

$$\frac{x}{C_1} - L_1\frac{\mathrm{d}I_1}{\mathrm{d}t} - M\frac{\mathrm{d}I_2}{\mathrm{d}t} = -\frac{y}{C_2} - L_2\frac{\mathrm{d}I_2}{\mathrm{d}t} - M\frac{\mathrm{d}I_1}{\mathrm{d}t} = 0$$

so that

$$x/C_1 + L_1\ddot{x} - M\ddot{y} = 0 \tag{10.22}$$

$$y/C_2 + L_2\ddot{y} - M\ddot{x} = 0 \tag{10.23}$$

using the notation $\dot{x} = \mathrm{d}x/\mathrm{d}t$. If both (10.22) and (10.23) are twice differentiated, yielding equations which we shall call (10.22a) and

(10.23a) respectively, then the $\ddot{y}$ and $\ddddot{y}$ in (10.23a) can be eliminated by using (10.22) and (10.22a). This gives, for x,

$$\ddddot{x}(1-k^2)+\ddot{x}(1/L_1C_1+1/L_2C_2)+x/L_1L_2C_1C_2 = 0 \quad (10.24)$$

where $k^2 = M^2/L_1L_2$, and the same equation for y. We expect the solution to be oscillatory rather than exponential and so we try $x = A\sin(\omega_N t+\phi)$ in (10.24):

$$\omega_N{}^4(l-k^2)-\omega_N{}^2(1/L_1C_1+1/L_2C_2)+1/L_1L_2C_1C_2 = 0 \quad (10.25)$$

For circuits with equal uncoupled natural frequencies, we can put $\omega_0 = 1/L_1C_1 = 1/L_2C_2$ from section 10.2 and therefore

$$\omega_N{}^4(1-k^2)-2\omega_0{}^2\omega_N{}^2+\omega_0{}^4 = 0$$

from which

$$\omega_N = \frac{\omega_0}{\sqrt{1\pm k}} \quad (10.26)$$

giving *two* natural frequencies separated by an amount depending on the tightness of coupling. The charges are given by

$$x = A\sin\left(\frac{\omega_0}{\sqrt{1+k}}t+\phi_A\right)+B\sin\left(\frac{\omega_0}{\sqrt{1-k}}t+\phi_B\right) \quad (10.27)$$

with 4 constants as demanded by the fourth order equation (10.24). If the solution for y is assumed to be of the form $C\sin(\omega_N t+\phi)$, then substitution in (10.22) gives $C = \mp(L_1/L_2)^{1/2}$, the upper sign corresponding to that of (10.26). Thus

$$y = -(L_1/L_2)^{1/2}\left[A\sin\left(\frac{\omega_0}{\sqrt{1+k}}t+\phi_A\right)-B\sin\left(\frac{\omega_0}{\sqrt{1-k}}t+\phi_B\right)\right] \quad (10.28)$$

In general, both x and y contain two terms which, for small k, have nearly equal frequencies so that beats are produced as in Fig. 10.5: the currents will be given by similar curves. It should be remembered that the presence of resistance will produce decaying oscillations and will slightly affect the frequencies.

The result is analogous to that for two coupled mechanical systems such as pendulums coupled through their suspensions, and these are treated in detail in textbooks on mechanics (see for example Becker, 1954). The two components of x and y are known as the *normal modes of oscillation* of the system. In giving only one

oscillator an initial displacement (corresponding to the charge Q_0 on C_1) both modes are excited and the motion is such that the energy initially associated completely with the first oscillator is transferred completely to the second, then back to the first and so on, as is evident from Fig. 10.5.

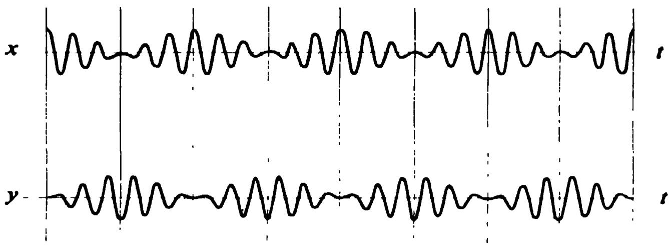


Fig. 10.5. Variation of charge in the mutually coupled circuits of Fig. 10.4.

The natural frequencies of other two-mesh networks can be obtained by similar methods (see problem 10.6).

10.4 General A.C. Theory

If we insert into the networks considered in the last two sections a sinusoidally varying e.m.f. and remove any D.C. sources, we now have a system with a continuous input of energy which is analogous to a mechanical system undergoing forced vibrations. The assumption that the steady current laws $\Sigma V=0$ and $\Sigma I=0$ apply at any instant will now yield differential equations differing from those of the last two sections only in the addition of the e.m.f.s: thus, if the circuit of Fig. 10.2b incorporates an alternating source $\mathscr{E}=\mathscr{E}_0 \sin \omega t$, the differential equation governing the charge Q is now

$$L\frac{\mathrm{d}^2Q}{\mathrm{d}t^2}+R\frac{\mathrm{d}Q}{\mathrm{d}t}+\frac{Q}{C}=\mathscr{E}_0 \sin \omega t \qquad (10.29)$$

and the current is given by $I=-\mathrm{d}Q/\mathrm{d}t$.

Equation (10.29) is of the form of (10.1) with $n=2$ and $\mathrm{f}(t)=\mathscr{E}_0 \sin \omega t$. The general solution of such an equation is the sum of the complementary function (CF) and a particular integral (PI). The CF is the general solution of the corresponding homogeneous equation, i.e. with $\mathrm{f}(t)=0$, and is thus identical with the transient solutions we have already dealt with. Because all these solutions

$\to 0$ as $t \to \infty$, we shall ignore the CF in future and consider only the PI which constitutes the *steady state* part of the solution. Calculation of PI's for simple circuits is not difficult (see e.g. Becker for the mechanical problem corresponding to (10.29)), but for complex networks we resort to methods which have analogies in the treatment of D.C. networks and which we shall develop in the remainder of this section.

Because differentiation of a sine or cosine generates terms with the same angular frequency, the linearity of the equations like (10.29) ensures that the variation of all currents and hence of potential differences occurs at the same frequency as that of the applied e.m.f. It follows that when the applied e.m.f. is $\mathscr{E} = \mathscr{E}_0 \sin \omega t$, all currents have the form $I_0 \sin(\omega t + \alpha)$ and all potential differences the form $V_0 \sin(\omega t + \beta)$. To apply the laws $\Sigma\ V = 0$ and $\Sigma\ I = 0$ we therefore need to know (*a*) the relations between V and I for the single circuit elements L, C, R and M, and (*b*) how to sum sinusoidally varying quantities of the same frequencies but of different phases. Operation (*a*) is dealt with below: operation (*b*) uses the phasor diagram or the complex representation described in appendix 10.i which the reader should consult at this stage.

Pure Resistance, Capacitance and Inductance. If we pass a current $I = I_0 \sin \omega t$ through either a pure R, a pure L or a pure C, the potential differences are obtained by using the relations of Fig. 10.1. We have for R

$V = RI = RI_0 \sin \omega t$

Hence $V = V_0 \sin \omega t$ where $V_0 = RI_0$.

For L,

$V = L\, \mathrm{d}I/\mathrm{d}t = \omega L I_0 \cos \omega t$

Hence $V = V_0 \sin(\omega t + \frac{1}{2}\pi)$ where $V_0 = \omega L I_0$.

For C,

$V = Q/C = \int I\, \mathrm{d}t/C = -(I_0 \cos \omega t)/\omega C$

Hence $V = V_0 \sin(\omega t - \frac{1}{2}\pi)$ where $V_0 = I_0/\omega C$.

These three important relations can be summarized by using either the phasor diagrams of Fig. 10.6 or a complex representation as follows:

In R, $V_0 = RI_0$ and V and I are in phase: $\mathbf{V} = R\mathbf{I}$

In L, $V_0 = \omega L I_0$ and V leads I by $\frac{1}{2}\pi$: $\mathbf{V} = j\omega L\mathbf{I}$

In C, $V_0 = I_0/\omega C$ and V lags on I by $\frac{1}{2}\pi$: $\mathbf{V} = \mathbf{I}/j\omega C$ or $-j\mathbf{I}/\omega C$ (10.30)

The quantities ωL and $1/\omega C$, known as *reactances*, play the same part as R in the relation between the *magnitudes* of V and I, but the phase differences must not be forgotten and the use of *complex reactances* $j\omega L$ and $-j/\omega C$ ensures this. A reactance in general will be denoted by X and a complex reactance by jX.

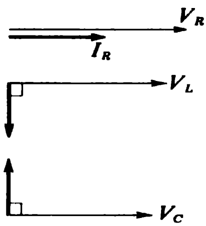


Fig. 10.6. Phase relations between current and potential difference in pure circuit elements.

For a mutual inductance M, the same treatment as above shows that a current $I = I_0 \sin \omega t$ in the primary produces a potential difference across the secondary such that $V_0 = \omega M I_0$ and V lags on I by $\frac{1}{2}\pi$. Hence $\mathbf{V} = -j\omega M\mathbf{I}$ (used in section 10.7 and subsequently).

Impedance. Any two-terminal network of R's, L's and C's will have a current of amplitude I_0, say, through it and a potential difference of amplitude V_0 across it as in Fig. 10.7a. The phase difference between the current and potential difference will in

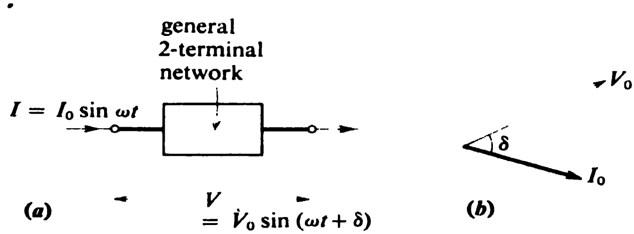


Fig. 10.7. (a) Current and potential difference for a general 2-terminal network; (b) phasor diagram for (a).

general be say δ, so that the network behaves neither as a pure resistance ($\delta = 0$) nor as a pure reactance ($\delta = \frac{1}{2}\pi$). However, by analogy with the relations for pure resistance and reactance above, we define the resistance of the network by

$$\frac{V_0 \cos \delta}{I_0} = R \text{ (Definition of } R) \qquad (10.31)$$

and the reactance of the network by

$$\frac{V_0 \sin \delta}{I_0} = X \quad \text{(Definition of } X\text{)} \qquad (10.32)$$

$V_0 \cos \delta$ being the component of V in phase with I, and $V_0 \sin \delta$ the component in quadrature. Using the complex representation, let $\mathbf{I} = I_0 e^{j\delta_1}$ and $\mathbf{V} = V_0 e^{j\delta_2}$ so that δ, the phase difference, is $(\delta_2 - \delta_1)$. Hence,

$$\mathbf{V} = V_0 e^{j\delta_2} = V_0 e^{j\delta} e^{j\delta_1} = (V_0 \cos \delta + jV_0 \sin \delta) e^{j\delta_1}$$

or

$$\mathbf{V} = (R + jX)\mathbf{I} \qquad (10.33)$$

using (10.31) and (10.32). The quantity $R+jX$ is the *complex impedance* $\mathbf{Z}$ of the two-terminal network so that

$$\mathbf{Z} = R + jX \quad \text{(Definition of } \mathbf{Z}\text{)} \qquad (10.34)$$

and

$$\mathbf{V} = \mathbf{ZI} \qquad (10.35)$$

The complex impedance contains the relations between the amplitudes and phases of current and potential difference for the network because from (10.31) and (10.32)

$$|Z| = (R^2 + X^2)^{1/2} = V_0/I_0; \qquad \arg Z = \tan^{-1}(X/R) = \delta \qquad (10.36)$$

$|Z|$ is often called simply the impedance of the network. Thus the response of a network is known if $\mathbf{Z}$ can be found in the form $R+jX$.

Generalized Network Theorems. The laws governing D.C. networks in chapter 6 were derived from the definition of resistance $V = RI$ together with the laws $\Sigma\, I = 0$ and $\Sigma\, V = 0$. A.C. networks are governed by $\mathbf{V} = \mathbf{ZI}$ coupled with $\Sigma\, \mathbf{I} = 0$ and $\Sigma\, \mathbf{V} = 0$, so that all the theorems of chapter 6 can be taken over with the substitution of *complex* potential differences, currents and impedances for V, I and R respectively.

Hence, for impedances in series

$$\mathbf{Z} = \mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_3 + \cdots = \Sigma\, \mathbf{Z}_i \qquad (10.37)$$

and in parallel

$$1/\mathbf{Z} = 1/\mathbf{Z}_1 + 1/\mathbf{Z}_2 + 1/\mathbf{Z}_3 + \cdots = \Sigma\, 1/\mathbf{Z}_i \qquad (10.38)$$

while the delta-Y transformation, Kirchhoff's laws, the superposition theorem and Thévenin's theorem all apply provided that $\mathbf{Z}$ is linear.

Simple Examples. From (10.30) the complex impedance of a

resistance is R, of an inductance is $j\omega L$ and of a capacitance is $-j/\omega C$ or $1/j\omega C$. Hence, for example,

L and R in series:

$$\mathbf{Z} = R+j\omega L, \qquad |Z| = (R^2+\omega^2L^2)^{1/2}, \qquad \delta = \tan^{-1}\omega L/R \tag{10.39}$$

C and R in parallel:

$$\mathbf{Z} = \mathbf{Z}_1\mathbf{Z}_2/(\mathbf{Z}_1+\mathbf{Z}_2) = -jR/\omega C(R-j/\omega C)$$

and hence $$|Z| = \left[\frac{R^2}{1+\omega^2C^2R^2}\right]^{1/2}, \qquad \delta = \tan^{-1}\omega CR \tag{10.40}$$

More complicated examples will be encountered later in the chapter. Phasor diagrams are best started with one of the potential differences or currents common to more than one component (see section 10.6).

Admittance and Susceptance. Just as a D.C. conductance can be defined as I/V for a two-terminal network, so we can define the conductance of the network of Fig. 10.7a by

$$\frac{I_0\cos\delta}{V_0} = G \quad \text{(Definition of } G\text{)} \tag{10.41}$$

and the *susceptance* by

$$\frac{-I_0\sin\delta}{V_0} = B \quad \text{(Definition of } B\text{)} \tag{10.42}$$

in which the current has been resolved instead of the potential difference as we did in (10.31) and (10.32). Without loss of generality, we take this time $\mathbf{V} = V_0$ and $\mathbf{I} = I_0e^{-j\delta}$ (the negative sign occurring here and in (10.42) merely because in our example $\mathbf{V}$ *leads* $\mathbf{I}$) and obtain

$$\mathbf{I} = I_0e^{-j\delta} = I_0\cos\delta - jI_0\sin\delta$$

or

$$\mathbf{I} = (G+jB)\mathbf{V} \tag{10.43}$$

The quantity $(G+jB)$ is the *complex admittance* $\mathbf{Y}$ so that

$$\mathbf{Y} = G+jB \quad \text{(Definition of } \mathbf{Y}\text{)} \tag{10.44}$$

and (10.43) can be written $\mathbf{I} = \mathbf{YV}$. Clearly also

$$\mathbf{Y} = 1/\mathbf{Z} \tag{10.45}$$

10.5 A.C. Power and R.M.S. Values

In a pure resistance R the instantaneous rate of production of heat is RI^2 no matter how I varies, and over a period of time the mean rate of heat production is $R\overline{I^2}$, where the bar indicates an average taken with respect to time. If we define a *root mean square* (R.M.S.) current by

$$I_{\mathrm{RMS}} = \sqrt{\overline{I^2}}$$

then the mean rate of heat production is RI^2_{RMS}. For a steady current, I_{RMS} and I are the same and so it follows that, as long as we agree to quote R.M.S. values, the heating effects of say 3 A D.C. and 3 A A.C. are the same. This applies equally to potential differences. An alternating e.m.f. quoted simply as 240 V is understood to be an R.M.S. value.

For a sinusoidally varying current $I = I_0 \sin \omega t$, I_0 is called the *peak current* and we have that

$$I^2_{\mathrm{RMS}} = \overline{I_0{}^2 \sin^2 \omega t} = \overline{\tfrac{1}{2}I_0{}^2 - \tfrac{1}{2}I_0{}^2 \cos 2\omega t} = \tfrac{1}{2}I_0{}^2$$

since the mean value of $\cos 2\omega t$ over a complete cycle or number of cycles is zero. Hence

$$I_{\mathrm{RMS}} = I_0/\sqrt{2} \tag{10.46}$$

for a sinusoidally varying current. It similarly follows that $V_{\mathrm{RMS}} = V_0/\sqrt{2}$. In section 10.4 we restricted ourselves to sinusoidal currents and potential differences and it follows that all the relations therein established involving V_0 and I_0 either as a ratio or in a scale diagram could equally well have used V_{RMS} and I_{RMS}.

Power in A.C. Circuits. The instantaneous rate of consumption of energy in a general 2-terminal network such as that of Fig. 10.7a is VI where $V = V_0 \sin(\omega t + \delta)$ and $I = I_0 \sin \omega t$. The power consumed is thus

$$\begin{aligned} P &= V_0 I_0 \sin \omega t \sin(\omega t + \delta) \\ &= V_0 I_0 \sin^2 \omega t \cos \delta + V_0 I_0 \sin \omega t \cos \omega t \sin \delta. \end{aligned}$$

The mean power over a cycle is thus

$$\bar{P} = \tfrac{1}{2}V_0 I_0 \cos \delta = \tfrac{1}{2}RI_0{}^2$$

because $\overline{\sin^2 \omega t} = \frac{1}{2}$ as above, $\overline{\sin \omega t \cos \omega t} = \overline{\frac{1}{2}\sin 2\omega t} = 0$, and where R is the real part of the complex impedance. Hence

$$\bar{P} = V_{\mathrm{RMS}} I_{\mathrm{RMS}} \cos \delta = RI_{\mathrm{RMS}}{}^2 \tag{10.47}$$

Similarly, the mean power consumption from a source of e.m.f. $\mathscr{E}_{\text{RMS}}$ supplying I_{RMS} at a phase difference δ is

$$\bar{P} = \mathscr{E}_{\text{RMS}} I_{\text{RMS}} \cos \delta \tag{10.48}$$

Cos δ is known as the *power factor* and is clearly 1 for a pure resistance and 0 for a reactance. Note that, because power is not a linear function of current or potential difference, it is not permissible to calculate it using the complex representation.

Impedance Matching. We now find the conditions for maximum power transfer to a load by a source of alternating e.m.f. with internal impedance (Fig. 10.8). The current I_0 is $\mathscr{E}_0/\{(R_1+R_2)^2+(X_1+X_2)^2\}^{1/2}$, and hence, from (10.47),

$$\bar{P} = R_2 I_{\text{RMS}}^2 = R_2 \mathscr{E}_{\text{RMS}}^2/\{(R_1+R_2)^2+(X_1+X_2)^2\} \tag{10.49}$$

If the load is given, then $\bar{P}$ is a maximum when $R_1=0$ and $X_1=-X_2$, but if, as is more common, the internal impedance is fixed

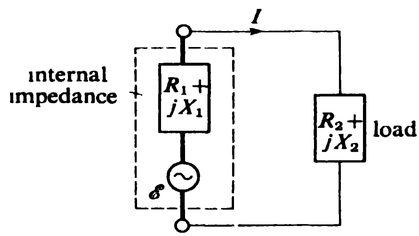


Fig. 10.8. Source with internal impedance and load.

then the power transfer depends on what parameters in the load can be chosen. If R_2 and X_2 are independent, differentiation of (10.49) with respect to R_2 shows that the optimum value of R_2 is $(R_1{}^2+(X_1+X_2)^2)^{1/2}$. If X_2 can be chosen as well, then $X_2=-X_1$ and $R_2=R_1$ are the conditions for maximum $\bar{P}$. The load is then *matched* to the source.

10.6 Resonance in Series and Parallel *LCR* Circuits

The differential equations governing A.C. circuits and forced vibrations in mechanical systems are identical in form and we should expect resonance phenomena to occur in one just as in the other.

Series Resonance. The series *LCR* circuit of Fig. 10.9a carries the same current through each of its components and this is the

starting phasor in the diagrams of Figs. 10.9b and c. In Fig. 10.9b the potential differences are drawn from the same point and compounded in the usual way to form the resultant, but Fig. 10.9c

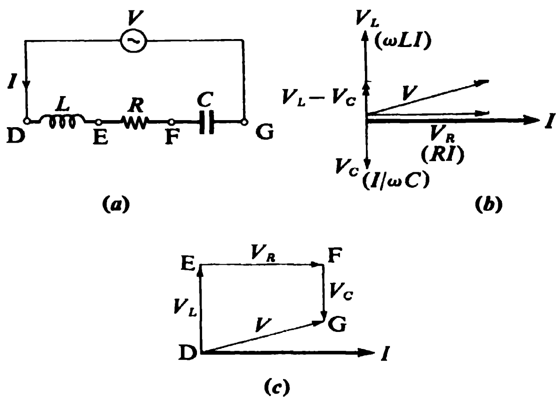


Fig. 10.9. (a) A series LCR circuit; (b) phasor diagram using a parallelogram: $V^2 = V_R^2 + (V_L - V_C)^2$ giving (10.50); (c) phasor diagram using a polygon.

illustrates a more useful method of arriving at the same result by using a polygon of phasors in which the corners can be labelled with the letters corresponding to junctions in the circuit.

The complex impedance of the combination is

$$\mathbf{Z} = R + j(\omega L - 1/\omega C)$$

and the magnitude of the current is thus

$$I = \frac{V}{\{R^2 + (\omega L - 1/\omega C)^2\}^{1/2}} \tag{10.50}$$

which can also be obtained from the phasor diagram. If either ω, L or C are varied, keeping all other quantities on the right-hand side of (10.50) constant, then the current variation shows a maximum when $\omega L = 1/\omega C$ and thus when **V** and **I** are *in phase*. Figure 10.10 shows the form of the curves when ω is varied for circuits with different R's. The occurrence of a maximum current at $\omega_0 = 1/(LC)^{1/2}$ (see section 10.2) is known as *current resonance* and corresponds to velocity resonance in a mechanical system. The potential difference across R is also a maximum at ω_0.

Corresponding to amplitude resonance is *charge resonance*. The charge circulating is I/ω and is thus obtainable from (10.50), curves

of Q against ω being given in Fig. 10.10b. The maximum is obtained as usual by differentiation and occurs at an angular frequency ω_Q given by

$$\omega_Q^2 = \omega_0^2 - R^2/2L^2 \tag{10.51}$$

It should be noted that neither charge nor current resonance occur at the natural frequency of the circuit given by (10.12) but that the

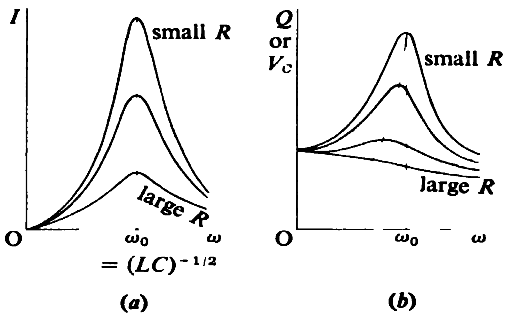


Fig. 10.10. Series LCR circuit. (a) Variation of current with frequency; (b) variation of charge with frequency.

three approach each other as R tends to zero. (See also problem 10.17.)

$\hat{Q}$ of a Circuit and Sharpness of Resonance. The quantity

$$2\pi \times \frac{\text{energy stored}}{\text{energy loss in one period}} \tag{10.52}$$

is known as the $\hat{Q}$ of a periodic system and for the series LCR circuit is $2\pi \times \frac{1}{2}LI_0^2/\frac{1}{2}RI_0^2T$ or $\omega L/R$. This varies with frequency and it is quite common to use the value at current resonance, $\omega_0 L/R$ or $\frac{1}{R}\sqrt{\frac{L}{C}}$, which we shall denote by $\hat{Q}_0$. Some previous results can be expressed anew: the critical damping condition in section 10.2, $L/CR^2 = \frac{1}{4}$, becomes $\hat{Q}_0 = \frac{1}{2}$; the logarithmic decrement from (10.20) and (10.21) is $2\pi(4\hat{Q}_0^2 - 1)^{-1/2}$ or simply $\pi/\hat{Q}_0$ if R is negligible; and the complex impedance of a series LCR circuit can be written as

$$\mathbf{Z} = R[1 + j\hat{Q}_0(\omega/\omega_0 - \omega_0/\omega)] \tag{10.53}$$

The sharpness of a resonance curve is defined by the width at

half-power when the current has a value $I_{max}/\sqrt{2}$, I_{max} being the resonant current V/R (Fig. 10.11). The width is thus $(\omega_1 - \omega_2)$ where ω_1 and ω_2 are the roots of

$$I_{max}/\sqrt{2} = V/R\sqrt{2} = V/[R^2 + (\omega L - 1/\omega C)^2]^{1/2}$$

using (10.50). These turn out to be $\pm R/2L \pm (R^2/4L^2 + 1/LC)^{1/2}$ of which those with the second negative sign are inadmissible because they give negative ω's. Hence

$$\delta\omega = (\omega_1 - \omega_2) = R/L = \omega_0/\hat{Q}_0 \qquad (10.54)$$

and a high-$\hat{Q}$ circuit is one with low damping ($\Lambda = \pi/\hat{Q}_0$) and sharp resonance.

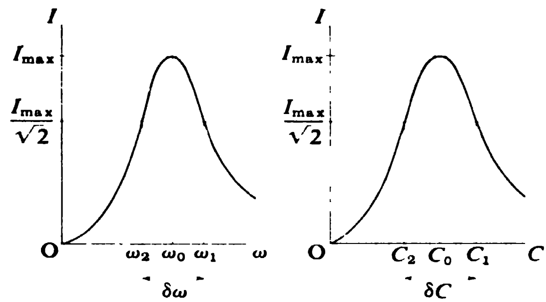


Fig. 10.11. Width of resonance curves.

If (10.50) is plotted for variation of C at a fixed frequency, a resonance curve of similar shape is obtained whose width can be shown to be $\delta C = 2C_0/\hat{Q}_0$ for high $\hat{Q}$. Measurement of $\hat{Q}_0$ is more easily achieved by using $2C_0/\delta C$ than by using (10.54) because of the greater ease of measuring small changes in capacitance.

Voltage Magnification. The maximum potential difference across R occurs at the resonant angular frequency ω_0 and is equal to the applied e.m.f. The maximum potential difference across C occurs at ω_Q, slightly lower than ω_0, while the maximum potential difference across L occurs at ω_L, slightly higher than ω_0 (see problem 10.17). At ω_0, the potential differences across L and C are opposite in phase but equal in magnitude to $\omega_0 L I_{max}$ or $\hat{Q}_0$ times the applied voltage. $\hat{Q}_0$ is thus the factor by which the applied voltage is magnified across L or C at resonance.

Parallel Resonance. The circuit of Fig. 10.12 has an impedance

$$\mathbf{Z} = \frac{(R+j\omega L)\times 1/j\omega C}{R+j(\omega L-1/\omega C)}$$

$$= \frac{\omega L - jR}{\omega CR + j(\omega^2 LC - 1)} \tag{10.55}$$

Hence
$$\frac{|V|}{|I|} = |Z| = \left(\frac{R^2+\omega^2 L^2}{\omega^2 C^2 R^2 + (\omega^2 LC-1)^2}\right)^{1/2} \tag{10.56}$$

which has a *maximum* value as the frequency varies. The current will thus have a *minimum* at a constant applied e.m.f. and the

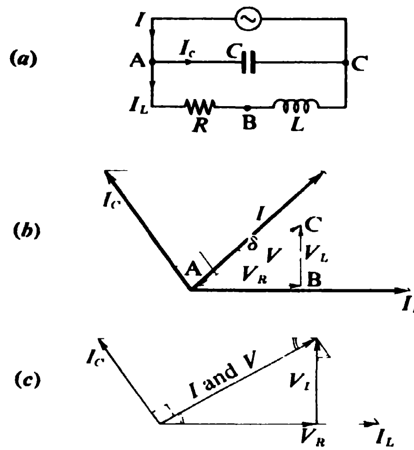


Fig. 10.12. (a) A parallel LCR circuit; (b) phasor diagram for (a); (c) phasor diagram at resonance (I and V in phase): by similar triangles $V_L/V_R = I_C/I$ or $\omega L/R = \omega CV/I$ and hence $V/I = L/CR$.

normal idea of resonance as a point of maximum response cannot apply. The definition of resonance normally accepted here is the frequency at which the current and potential difference are in phase. This means that the reactive part of the impedance must be zero and from (10.55) this occurs when the angular frequency is

$$\omega_P^{\,2} = \omega_0^{\,2} - R^2/L^2 \tag{10.57}$$

where $\omega_0^{\,2} = 1/LC$. At ω_P, the impedance given by (10.56) is L/CR, known as the dynamic resistance of the network. Once again resonance only occurs at ω_0 for negligible damping.

The $\hat{Q}$ of the circuit will be $\omega L/R$ by (10.52) and $\hat{Q}_0$ will again be taken as $\omega_0 L/R$. At resonance the phasor diagram of Fig. 10.12c shows that the phase angle between I_C and I_L is $\frac{1}{2}\pi+\delta$ where $\delta=\tan^{-1}\omega_0 L/R=\tan^{-1}\hat{Q}_0$. Thus for a high-$\hat{Q}$ circuit, I_L and I_C will be nearly π out of phase, equal in magnitude and each will be in quadrature with the main current. The magnitude of both will be $\omega_0 CV$ while the main current is V/dynamic resistance or VCR/L. The ratio of I_C or I_L to the main current at resonance is thus $\omega_0 L/R$ or $\hat{Q}_0$, which is again a magnification factor but this time of current.

Use of Resonant Circuits. The series and parallel circuits are complementary: the sharp maximum response of a series circuit makes it suitable for picking out a small band of frequencies from a wide range and in this respect it acts as a filter. The parallel circuit, on the other hand, has a sharp maximum $|Z|$ at resonance (Fig. 10.13) and where the amplification of a valve circuit depends

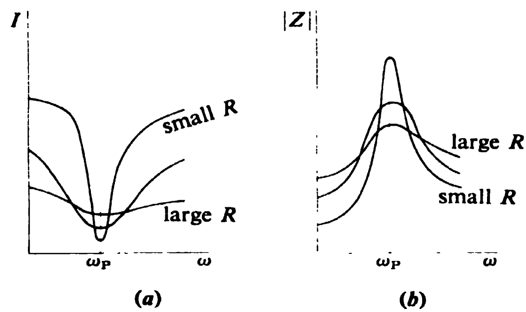


Fig. 10.13. Variation of (a) current and (b) impedance in the parallel LCR circuit of Fig. 10.12a.

on an impedance the parallel tuned circuit gives high amplification over only a small range of frequencies.

10.7 Coupled Circuits and the Ideal Transformer

The network of Fig. 10.14 is fundamental to this and the next sections. It consists of two meshes: the primary, containing a source $\mathbf{V}$ and carrying a mesh current $\mathbf{I}_p$, and the secondary, with no source and a mesh current $\mathbf{I}_s$. In this section we shall work mainly in terms of the total primary impedance $\mathbf{Z}_p$ $(=\mathbf{Z}_a+\mathbf{Z}_m)$ and the total secondary impedance $\mathbf{Z}_s$ $(=\mathbf{Z}_b+\mathbf{Z}_m+\mathbf{Z}_L)$ where $\mathbf{Z}_m$ is mutual and $\mathbf{Z}_L$ is a load.

The impedance presented to $\mathbf{V}$ is $\mathbf{Z}_a + \mathbf{Z}_m(\mathbf{Z}_b + \mathbf{Z}_L)/(\mathbf{Z}_m + \mathbf{Z}_b + \mathbf{Z}_L)$ and by using $\mathbf{Z}_p$ and $\mathbf{Z}_s$ we obtain

$$\mathbf{Z}_{in} = \mathbf{Z}_p - \frac{\mathbf{Z}_m{}^2}{\mathbf{Z}_s} \tag{10.58}$$

and hence

$$\mathbf{I}_p = \frac{\mathbf{V}\mathbf{Z}_s}{\mathbf{Z}_p\mathbf{Z}_s - \mathbf{Z}_m{}^2} \tag{10.59}$$

The secondary current can be obtained by Thévenin's theorem: the whole of the network to the left of $\mathbf{Z}_L$ can be replaced by a generator

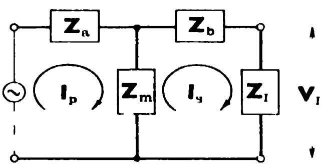


Fig. 10.14. General 2-mesh network: $\mathbf{Z}_p = \mathbf{Z}_a + \mathbf{Z}_m$; $\mathbf{Z}_s = \mathbf{Z}_b + \mathbf{Z}_m + \mathbf{Z}_l$.

of e.m.f. $\mathbf{V}\mathbf{Z}_m/\mathbf{Z}_p$ in series with an impedance $\mathbf{Z}_b + \mathbf{Z}_m\mathbf{Z}_a/(\mathbf{Z}_m + \mathbf{Z}_a)$ which can be written as $\mathbf{Z}_s - \mathbf{Z}_m{}^2/\mathbf{Z}_p$. Hence

$$\mathbf{I}_s = \frac{\mathbf{V}\mathbf{Z}_m}{\mathbf{Z}_p\mathbf{Z}_s - \mathbf{Z}_m{}^2} \tag{10.60}$$

while the potential difference across the load is

$$\mathbf{V}_L = \mathbf{Z}_L\mathbf{I}_s = \frac{\mathbf{V}\mathbf{Z}_m\mathbf{Z}_L}{\mathbf{Z}_p\mathbf{Z}_s - \mathbf{Z}_m{}^2} \tag{10.61}$$

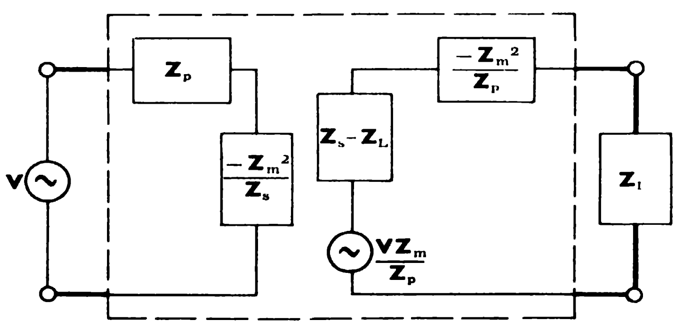


Fig. 10.15. Equivalent circuits for Fig. 10.14 as seen from generator and load.

Figure 10.15 shows how equations (10.59) and (10.60) can be interpreted to give equivalent circuits as seen from the generator and load respectively.

Mutual Inductance Coupling: No Resonance. We first consider coupling by a mutual inductance M between circuits with negligible capacitance so that resonance does not occur at low frequencies. By the transformation of equations (9.43), the actual circuit of Fig. 10.16a may be replaced by that of Fig. 10.16b, as may be checked by

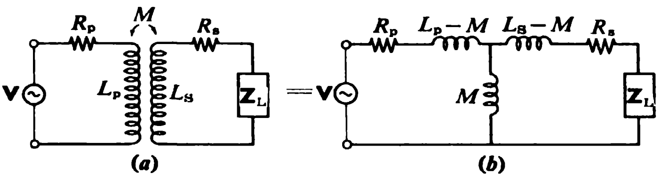


Fig. 10.16. Non-resonant circuits coupled by a mutual inductance. R_s and L_S are the resistance and inductance of the secondary winding.

writing down the equations for the two sets of meshes. The equations derived above will apply with $\mathbf{Z}_m = -j\omega M$, $\mathbf{Z}_p = R_p + j\omega L_p$, $\mathbf{Z}_s = R_s + jX_s$ so from (10.58)

$$\mathbf{Z}_{in} = \left(R_p + \frac{\omega^2 M^2 R_s}{|Z_s|^2}\right) + j\left(\omega L_p - \frac{\omega^2 M^2 X_s}{|Z_s|^2}\right) \tag{10.62}$$

showing an effective increase in primary resistance and decrease in primary inductance due to the presence of the secondary. R_s and X_s include both the load and the rest of the secondary circuit.

An *ideal transformer* (Fig. 10.17) is one with complete transference of energy from generator to load (i.e. no losses), and with

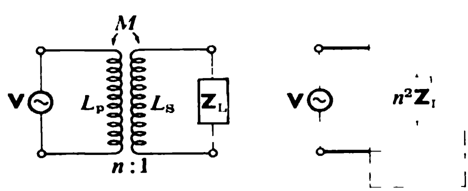


Fig. 10.17. Ideal transformer ($R_p = R_s = 0$; $M = (L_p L_s)^{1/2}$) and the reflected impedance as it appears to the generator.

no flux leakage. The former property means that no effective resistances occur except in the load, and the latter that if N_p and N_s

are the numbers of turns on the primary and secondary windings

$$L_p = bN_p{}^2; \qquad L_S = bN_s{}^2 \quad \text{by equation (9.21)}$$

$$M = \sqrt{L_p L_S} = bN_p N_s \quad \text{because } k = 1$$

where b is a constant and we distinguish carefully between L_S, the inductance of the winding, and L_s, the total secondary inductance. The ratio $\mathbf{V}_L/\mathbf{V}$ is, by (10.61)

$$\frac{\mathbf{V}_L}{\mathbf{V}} = \frac{-j\omega M \mathbf{Z}_L}{j\omega L_p(j\omega L_S + \mathbf{Z}_L) + \omega^2 M^2} = -\frac{M}{L_p}$$

or
$$\frac{\mathbf{V}_L}{\mathbf{V}} = -\frac{N_s}{N_p} = -\frac{1}{n} \tag{10.63}$$

n being the turns ratio. This could also have been seen by realizing that, if Φ is the magnetic flux per turn, $\mathbf{V} = +N_p\,\mathrm{d}\Phi/\mathrm{d}t$ and $\mathbf{V}_L = -N_s\,\mathrm{d}\Phi/\mathrm{d}t$. Equation (10.63) shows that the potential difference is stepped up or down in the inverse ratio of n and that the potential differences are π out of phase. By (10.59) and (10.60)

$$\frac{\mathbf{I}_s}{\mathbf{I}_p} = \frac{\mathbf{Z}_m}{\mathbf{Z}_s} = -\frac{j\omega M}{j\omega L_S + \mathbf{Z}_L} \tag{10.64}$$

If L_S is very large, $\mathbf{Z}_L$ can be neglected and $\mathbf{I}_s/\mathbf{I}_p$ is approximately $-M/L_S$ or $-n$. The current is thus stepped in the inverse ratio to the potential difference. In this case the load as seen by the generator is $\mathbf{V}/\mathbf{I}_p = n^2\mathbf{V}_L/\mathbf{I}_s$ or $n^2\mathbf{Z}_L$ so that for an ideal transformer

$$\mathbf{Z}_{in} \doteqdot n^2 \mathbf{Z}_L \tag{10.65}$$

known as the *reflected impedance*. This can be confirmed by using (10.62) which shows that L_p and the effect of L_S cancel, leaving only the effect of the load inductance in the reactive part and the load resistance in the resistive part ($R_p = 0$ because no losses occur). The property (10.65) is important because it enables impedance matching to be achieved by a proper choice of n. (See next section for real transformers.)

Mutual Inductance Coupling: Resonance. If both primary and secondary circuits incorporate capacitance then resonance in possible and at least for frequencies near resonances the resistances cannot be neglected as they were above. Moreover, we shall now introduce the possibility of leakage by varying the coupling coefficient k $(= M/\sqrt{L_p L_s}$ where L_s is still the total secondary in-

ductance). Identical primary and secondary circuits show the main features and simplify the algebra: we shall therefore adopt $\mathbf{Z}_m = -j\omega M$, $\mathbf{Z}_p = \mathbf{Z}_s = R + j(\omega L - 1/\omega C)$, $M = kL$ as in Fig. 10.18. We consider in turn the currents *at* the angular frequency $\omega_0 = 1/\sqrt{LC}$, *near* ω_0 and finally *remote from* ω_0.

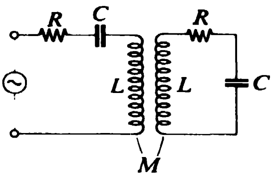


Fig. 10.18. Identical resonant circuits coupled by a mutual inductance.

At ω_0, $\mathbf{Z}_p = \mathbf{Z}_s = R$ and from (10.59) and (10.60)

$$\mathbf{I}_p = \frac{\mathbf{V}R}{R^2 + \omega_0{}^2 k^2 L^2}; \qquad \mathbf{I}_s = \frac{-j\omega_0 kL\mathbf{V}}{R^2 + \omega_0{}^2 k^2 L^2} \tag{10.66}$$

These show that $\mathbf{I}_p$ at ω_0 increases continually as the coupling becomes looser while $\mathbf{I}_s$ first increases and then falls, its maximum occurring when $k = R/\omega_0 L = 1/\hat{Q}_0$.

Near ω_0, let $\omega - \omega_0 = \mathrm{d}\omega$, when $(\omega L - 1/\omega C)$ or $\omega_0 L(\omega/\omega_0 - \omega_0/\omega)$ can be written to a high accuracy as $2L\,\mathrm{d}\omega$. It follows that near ω_0

$$\mathbf{I}_p = \frac{\mathbf{V}(R + j2L\,\mathrm{d}\omega)}{(R + jL\,\mathrm{d}\omega)^2 + \omega_0{}^2 k^2 L^2}; \qquad \mathbf{I}_s = \frac{-j\omega_0 kL\mathbf{V}}{(R + j2L\,\mathrm{d}\omega)^2 + \omega_0{}^2 k^2 L^2} \tag{10.67}$$

Treating $\mathrm{d}\omega$ as the variable, we can obtain the maxima and minima by differentiating and equating to zero the moduli of $\mathbf{I}_p$ and $\mathbf{I}_s$. For $\mathbf{I}_s$, only the denominator contains the variable and its modulus is

$$[R^2 - 4L^2(\mathrm{d}\omega)^2 + \omega_0{}^2 k^2 L^2]^2 + 16L^2(\mathrm{d}\omega)^2 R^2$$

Thus for maxima and minima of $|I_s|$

$$4L^2(\mathrm{d}\omega)^3 + (R^2 - \omega_0{}^2 k^2 L^2)\,\mathrm{d}\omega = 0$$

The three roots of this equation can be written

$$\mathrm{d}\omega = 0 \quad \text{or} \quad \pm\frac{\omega_0}{2\hat{Q}_0}\sqrt{(\hat{Q}_0 k)^2 - 1}$$

Thus if $\hat{Q}_0 k > 1$ there are 3 real roots, $d\omega = 0$ being a minimum and the other two being maxima, while if $\hat{Q}_0 k < 1$ there is only one real root at $\omega = \omega_0$, which is now a maximum. $k_c = 1/\hat{Q}_0$ is critical coupling where 3 roots coincide, tighter coupling giving two peaks, looser one (Fig. 10.19). Since k cannot be greater than 1, $\hat{Q}_0$ must also be at least 1 if a double peak is to occur. The variations in $|I_p|$ are similar.

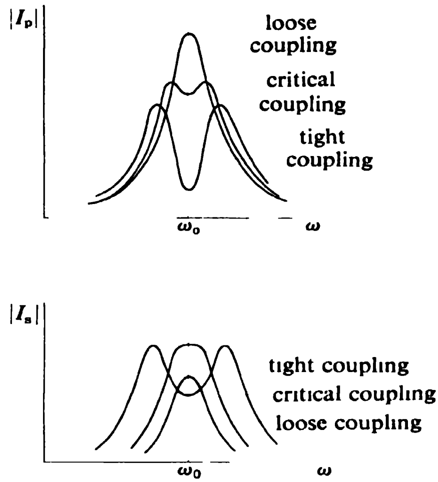


Fig. 10.19. Primary and secondary currents in the coupled circuits of Fig. 10.18.

With tight coupling the peaks are remote from ω_0 and the resistances can usually be neglected. This results in the denominators of $\mathbf{I}_p$ and $\mathbf{I}_s$ both being given by $\omega^2 k^2 L^2 - (\omega L - 1/\omega C)^2$ which has a minimum value of zero when

$$\omega = \sqrt{\frac{\omega_0}{1 \pm k}} \tag{10.68}$$

the same frequencies as those of (10.26).

Direct Coupling: Resonant Circuits. Circuits may be directly coupled through a self-inductance or a capacitance as in Fig. 10.20. The definition of k is now generalized to $X_m/\sqrt{X_p X_s}$ where X_m is the mutual reactance and X_p and X_s are the total reactances *of the same kind as* X_m in the primary and secondary. In Fig. 10.20 for example, $X_m = 1/\omega C_m$ and $X_p = X_s = 1/\omega C$ so that $k = C/C_m$.

Neglecting resistance, the circuits will have maximum currents when $\mathbf{Z}_p\mathbf{Z}_s - \mathbf{Z}_m{}^2$ is zero, i.e. when

$$[j(\omega L - 1/\omega C)]^2 - (1/j\omega C_m)^2 = 0$$

or

$$\omega_0{}^4\omega^4 - 2\omega_0{}^2\omega^2 + 1 - C^2/C_m{}^2 = 0$$

where $\omega_0 = 1/\sqrt{LC}$. Hence the currents again have double maxima at

$$\omega = \frac{\omega_0}{\sqrt{1 \pm k}}$$

The properties of coupled circuits of all types are broadly similar: loose coupling means that the circuits are virtually independent and

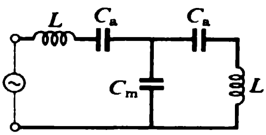


***Fig. 10.20.** Direct-coupled resonant circuits. The total primary capacitance C_p and secondary capacitance C_s are $C_p = C_s = C_aC_m/(C_a + C_m)$. If $C_p = C_s = C$, then $k = C/C_m$.*

resonance occurs at one frequency in the primary and one in the secondary (coincident in Fig. 10.19). Tighter coupling produces a double resonance in both circuits at frequencies more widely separated as k increases. For near-critical coupling the response is broad at the peak but drops on both sides more sharply than that for a single resonant circuit: such a network can be used to pass only a band of frequencies with nearly constant amplification.

More detailed analysis shows that when the primary and secondary resonate at the same frequency but have unequal $\hat{Q}_0$'s, $\hat{Q}_p$ and $\hat{Q}_s$, critical coupling is given by $k_c = 1/\sqrt{\hat{Q}_p\hat{Q}_s}$. If the primary is non-resonant and the secondary resonant, the currents show maxima which are shifted from ω_0: this can be seen to follow from (10.59) for $\mathbf{I}_p$. Finally, if the resonant frequencies of primary and secondary are unequal, the two peaks obtained at tight coupling are of unequal height.

10.8 Transformers in Practice

Energy losses in real transformers result from joule heating in the windings (copper losses) and in the core (eddy current losses)

together with hysteresis loss (section 14.6) also in the core. Copper losses are equivalent to resistances r_p and r_s in Fig. 10.21 while the core or iron losses are represented by a shunt r_c. The flux leakage which makes k less than 1 can be represented by splitting off $(1-k)L_p$ from the inductance of the primary winding so that only kL_p takes the current which produces the magnetic flux in the core.

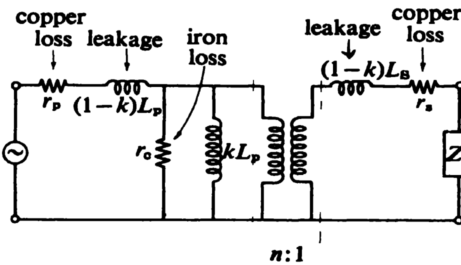


Fig. 10.21. Equivalent circuit of transformer with losses. The ideal transformer in the dotted rectangle is meant to indicate that the effect of the secondary on the primary is that of $n^2 \times$ the impedances on the right, in parallel with kL_p.

The effect of the secondary is that of leakage inductance, copper losses r_s and the load reflected back into the primary as if through an ideal transformer with a turns ratio $n:1$: the reader should be able to show from (10.62) that this is so. There will in addition be capacitance associated with the primary and secondary.

Power Transformers. For reducing or increasing a voltage at power frequencies, usually 50 c/s, a transformer will be wound on a ferromagnetic core which reduces leakage and makes k nearly 1. The resultant iron losses are reduced by lamination (minimizing eddy currents) and by choosing a material with a narrow hysteresis loop—silicon steel is commonly used. At low frequencies the effect of self-capacitance is also negligible and the transformer approximates to an ideal one (see Fig. 10.22a for construction).

An auto-transformer is shown in Fig. 10.22b. It is used when the isolation of primary from secondary for D.C. potentials is not important and has the advantages of reducing leakage of flux still further and copper losses as well. The 'variac' of Fig. 10.22c is usually employed when continuous variation of alternating potential difference is required ranging from zero to the maximum available.

Audio-frequency and Radio-frequency Transformers. Over the

range from about 200 c/s to 20 kc/s transformers are used mainly in amplifiers to couple one stage to the next without passing on a D.C. potential, or for matching the output to a device such as a loudspeaker. At the lower end of the range the same considerations apply as to the power transformer but at the upper end the possibility of resonance occurs because of the self-capacitances,

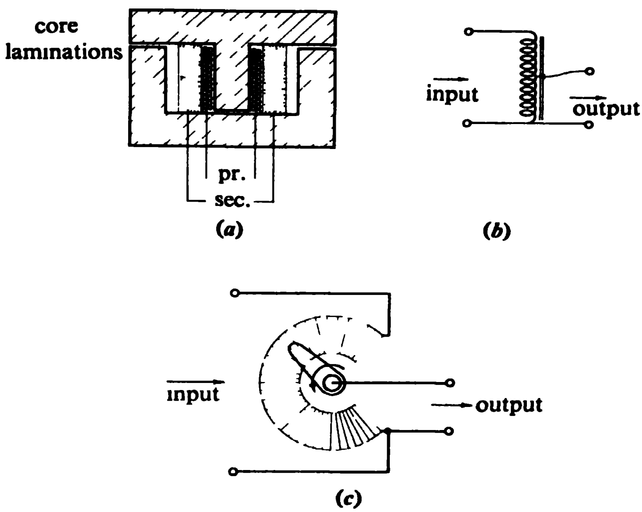


Fig. 10.22. (a) Construction of a small power transformer; (b) an auto-transformer; (c) a 'variac'.

whereas a uniform response over the range is desirable. Low frequency response can be improved by increasing the incremental inductance (section 14 6).

At radio-frequencies, from about 100 kc/s to 10 Mc/s, the properties of coupled resonant circuits are desirable so that narrow bands shall be passed with a sharp cut-off outside them. R-F transformers are thus invariably tuned. The losses in normal ferromagnetic cores are large at these high frequencies and either air or dust cores are used (section 14.6).

Pulse Transformers. Since a train of pulses contains a very wide range of frequencies, its transmission by a transformer without appreciable distortion presents great difficulties. Cores must have a constant permeability up to high frequencies as well as high resistivity so that ferrites (section 14.1) are commonly used.

10.9 Filters and Attenuators

Networks with a pair each of input and output terminals and containing no sources of e.m.f. are sometimes known as *passive quadripoles* although we shall refer to them simply as 4-terminal networks, the transformer being a typical example. If the ratio of output to input is less than 1 and independent of frequency the network is an *attenuator*—the potential divider of Fig. 6.17b is an example. If the ratio of output to input is almost 1 for a range of frequencies but falls off sharply outside this range, the network is a *filter*—the coupled resonant circuits of section 10.7 are examples.

Both filters and attenuators become more versatile if they consist of more than one stage and a common form is the symmetrical ladder network shown in Fig. 10.23 which can be divided into a

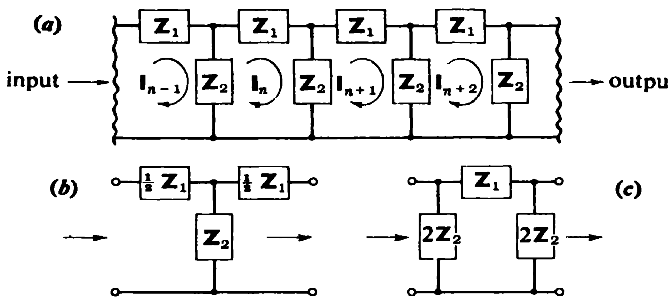


Fig. 10.23. (a) Symmetrical ladder network which can be built up from (b) identical T-sections by splitting $\mathbf{Z}_1$ or (c) identical π-sections by splitting $\mathbf{Z}_2$.

number of identical T- or π-sections as shown (these can be identified with Y- and Δ-networks). An important property of a 4-terminal network is its *characteristic impedance*, defined as that load $\mathbf{Z}_k$ which, connected to the output terminals, makes the input impedance also $\mathbf{Z}_k$: any number of identical networks connected output to input will still have the same characteristic impedance.

As an example, the T-section of Fig. 10.23b has

$$\mathbf{Z}_k = \tfrac{1}{2}\mathbf{Z}_1 + \mathbf{Z}_2 - \frac{\mathbf{Z}_2^{\,2}}{\mathbf{Z}_2 + \frac{1}{2}\mathbf{Z}_1 + \mathbf{Z}_k}$$

from (10.58) with $\mathbf{Z}_a = \mathbf{Z}_b = \frac{1}{2}\mathbf{Z}_1$ and $\mathbf{Z}_m = \mathbf{Z}_2$. Simplifying,

$$\mathbf{Z}_{kT} = \sqrt{\mathbf{Z}_1\mathbf{Z}_2 + \tfrac{1}{4}\mathbf{Z}_1^{\,2}} \tag{10.69}$$

For the π-section of Fig. 10.23c, convert it to its equivalent T-section by a Λ-Y transformation and use (10.69) when

$$\mathbf{Z}_{k\pi} = \mathbf{Z}_1\mathbf{Z}_2\sqrt{\frac{1}{\mathbf{Z}_1\mathbf{Z}_2+\frac{1}{4}\mathbf{Z}_1^{\,2}}} \tag{10.70}$$

and thus $\mathbf{Z}_{kT}\mathbf{Z}_{k\pi}=\mathbf{Z}_1\mathbf{Z}_2$. We shall work wholly in terms of T-sections.

If the load is in fact $\mathbf{Z}_k$, the network is said to be correctly terminated and it should be clear that if this is the case the load on a generator at the input is equal to $\mathbf{Z}_k$ no matter how many sections are used. $\mathbf{Z}_k$ is also known as the *iterative impedance* for ladder networks.

The output:input ratio which determines the filtering and attenuating is expressed in terms of a *propagation constant* γ defined by

$$\frac{\mathbf{I}_n}{\mathbf{I}_{n+1}} = e^{\gamma} \tag{10.71}$$

where $\mathbf{I}_n$ and $\mathbf{I}_{n+1}$ are the mesh currents in successive loops of a correctly terminated network (Fig. 10.23a). The point of this is that for n identical sections the propagation constant will be n times that for one. For a single T-section, equations (10.59) and (10.60) give

$$\frac{\mathbf{I}_p}{\mathbf{I}_s} = \frac{\mathbf{Z}_s}{\mathbf{Z}_m} = \frac{\mathbf{Z}_{kT}+\frac{1}{2}\mathbf{Z}_1+\mathbf{Z}_2}{\mathbf{Z}_2}$$

and so from (10.69) and putting $\mathbf{Z}_1/2\mathbf{Z}_2=r$,

$$e^{\gamma} = 1+r+\sqrt{r^2+2r} \tag{10.72}$$

In general γ will be complex so let

$$\gamma = \alpha+j\beta; \qquad e^{\gamma} = e^{\alpha}\cos\beta+je^{\alpha}\sin\beta \tag{10.73}$$

Restricting r to real values, equation (10.72) shows that if r^2+2r is $\geqslant 0$ then $e^{\alpha}\sin\beta=0$ and $e^{\alpha}\cos\beta$ is finite so that $\beta=0$ and γ is real. On the other hand, if $r^2+2r<0$ then $e^{\alpha}\cos\beta=(1+r)$ and $e^{\alpha}\sin\beta=\sqrt{-(r^2+2r)}$ whence $e^{2\alpha}=1$ so that $\alpha=0$ and γ is imaginary. Thus only when $r^2+2r<0$ is there no attenuation and this is only true when r lies between 0 and -2. To summarize:

$-2 < r < 0$:	no attenuation, $\mathbf{I}$ changes phase by amount depending on r
$r \leqslant -2;\ r \geqslant 0$:	attenuation of amount dependent on r, phase change of 0 or π.

Filters. A ladder network thus passes without attenuation frequencies for which $\mathbf{Z}_1/4\mathbf{Z}_2$ lies between 0 and -1 and attenuates the rest. Filters thus require $\mathbf{Z}_1$ and $\mathbf{Z}_2$ to be reactances of opposite sign. If $\mathbf{Z}_1 = j\omega L$ and $\mathbf{Z}_2 = 1/j\omega C$ then frequencies are passed for which $\frac{1}{4}\omega^2 LC$ lies between 0 and 1, i.e. all frequencies up to $\omega = 2/\sqrt{LC}$ are passed and Fig. 10.24a shows a low-pass filter.

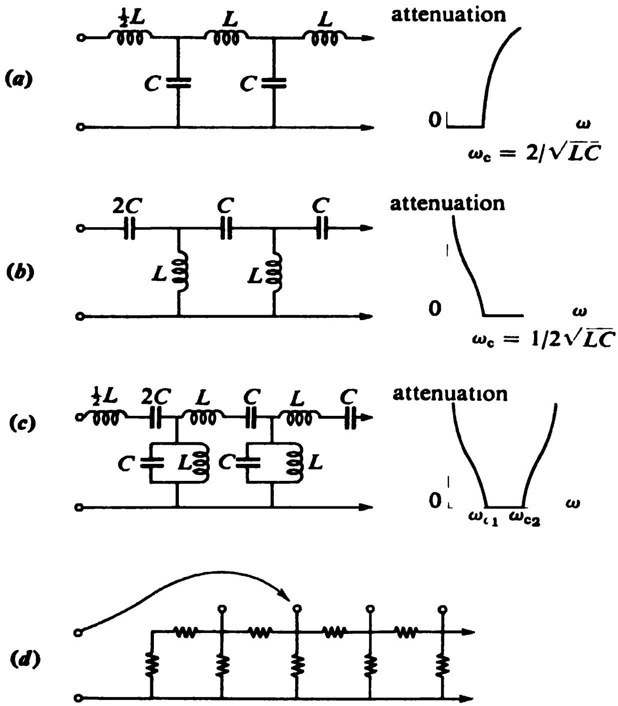


Fig. 10.24. (a) Low-pass filter; (b) high-pass filter; (c) band-pass filter; (d) ladder attenuator.

The other results of the figure are easily calculated: the shape of the attenuation curves can be obtained from the variation of α with frequency (problem 10.26).

Sections with the same $\mathbf{Z}_k$ but with different $\mathbf{Z}$'s can be connected to obtain desirable characteristics in the pass band or in the attenuation curve and if the load is not equal to $\mathbf{Z}_k$, terminal half-

sections (problem 10.23) can be used for matching. Note that a low-pass filter is used as a smoothing circuit after rectification of A.C. in D.C. power supplies.

Attenuators. If $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are both resistances, r is always real and positive, and attenuation therefore takes place without change of phase. A ladder network is superior to a simple potential divider because with proper termination the load on the generator is independent of the number of sections and therefore of the attenuation: this is not the case with a potential divider unless it feeds a system with a very high input impedance such as that between grid and cathode of a valve (see problem 6.20).

The ratio of powers P_2 to P_1 is expressed in *bels* by $\log_{10} P_2/P_1$, though more commonly the *decibel* (dB) is used and a gain in dB is thus $10 \log_{10} P_2/P_1$. For a correctly terminated ladder network of resistances the ratio (output power: input power) is $V_L I_s/V I_p$ using the notation of Fig. 10.14 but, since V sees a resistance equal to the load, $V/I_p = V_L/I_s$ and the power gain is thus $20 \log_{10} I_s/I_p$ or $-20 \log_{10} e^{\gamma}$ by (10.71). Hence because $\gamma = \alpha$ here,

$$\text{Loss in attenuator} = 20\alpha \log_{10} e \quad \text{dB per section}$$

10.10 Transmission Lines

Electric power is often carried over a distance by cables consisting of two conductors which may be thought of as conveying current to and from the generator. We have seen in sections 5.2 and 9.6 that such cables will possess distributed capacitance and self-inductance, and we see also that they will have resistance along the conductors and conductance between them. If C, L, R and G are the magnitudes of these parameters per unit length of the cable (R for both conductors) then a small section of length dx can be represented diagrammatically as in Fig. 10.25a, x being the distance measured from left to right. The complete line is thus a ladder network in which the impedances are continuous instead of lumped and we might expect to find similarities between these lines and filters.

Let the complex current and potential difference at the input end of the elementary section be $\mathbf{I}$ and $\mathbf{V}$, when the change $d\mathbf{V}$ is a result of a fall in potential along the impedance $(R + j\omega L)\,dx$. If we let $\mathbf{Z} = R + j\omega L$,

$$\mathbf{Z}\,dx\,\mathbf{I} = -d\mathbf{V}$$

or

$$-\frac{\partial \mathbf{V}}{\partial x} = \mathbf{Z}\mathbf{I} \tag{10.74}$$

The increment in current $d\mathbf{I}$ is a result of that bypassed by the admittance $\mathbf{Y}\,dx = (G + j\omega C)\,dx$ so that

$$-\frac{\partial \mathbf{I}}{\partial x} = \mathbf{YV} \tag{10.75}$$

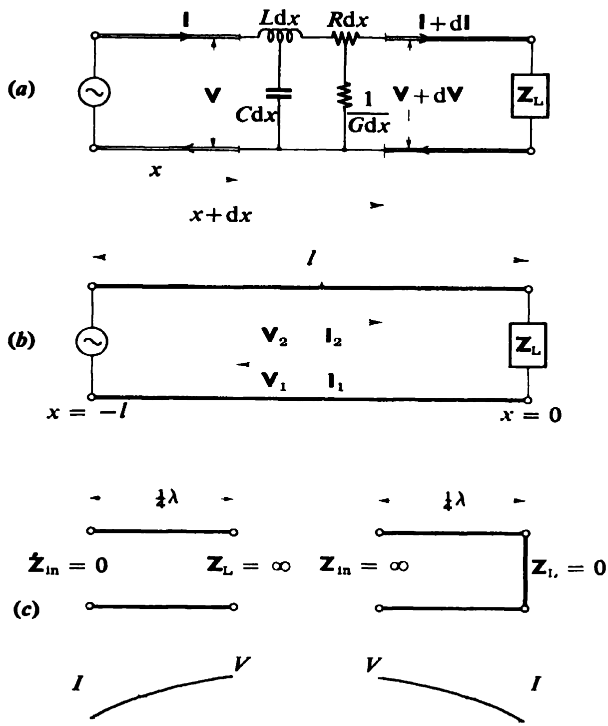


Fig. 10.25. (a) Transmission line parameters; (b) reflection from mismatched line; (c) voltage and current variations in $\frac{1}{4}\lambda$ lines.

the partial derivatives being used because variations with time occur as well. It follows that

$$\frac{\partial^2 \mathbf{I}}{\partial x^2} = \mathbf{YZI}; \qquad \frac{\partial^2 \mathbf{V}}{\partial x^2} = \mathbf{YZV} \tag{10.76}$$

and hence

$$\mathbf{I} = \mathbf{I}_1 e^{\gamma x} + \mathbf{I}_2 e^{-\gamma x}; \qquad \mathbf{V} = \mathbf{V}_1 e^{\gamma x} + \mathbf{V}_2 e^{-\gamma x} \tag{10.77}$$

where $\gamma = \sqrt{\mathbf{YZ}}$ *, called the propagation constant as in the last section. In general γ is complex so let

$$\gamma = \alpha + j\beta \tag{10.78}$$

The physical meaning of these terms is brought out when we remember that the instantaneous values of current and potential difference are given by the real or imaginary parts of $\mathbf{I}e^{j\omega t}$ and $\mathbf{V}e^{j\omega t}$ and we have

$$\mathbf{I}e^{j\omega t} = \mathbf{I}_1 e^{\alpha x} e^{j(\omega t + \beta x)} + \mathbf{I}_2 e^{-\alpha x} e^{j(\omega t - \beta x)} \tag{10.79}$$

with a similar expression for $\mathbf{V}$. The second term is an attenuated wave moving from left to right, the first a similar wave travelling from right to left. α is thus an attenuation constant and β a phase constant determining the wave velocity, which is ω/β (and therefore $\beta = 2\pi/\lambda$, λ being the wave-length). The reader should confirm that this is so by considering the variation of each term, first at constant x and varying t, and then at constant t and varying x.

Because α and β are in general frequency-dependent, any input waveform containing a range of frequencies is distorted and dispersion occurs.

Characteristic Impedance. Returning to the solution (10.77), we find by substitution into (10.75) that

$$-\gamma \mathbf{I}_1 e^{\gamma x} + \gamma \mathbf{I}_2 e^{-\gamma x} = \mathbf{YV}_1 e^{\gamma x} + \mathbf{YV}_2 e^{-\gamma x}$$

for *all* x and hence, equating coefficients,

$$-\gamma \mathbf{I}_1 = \mathbf{YV}_1; \qquad \gamma \mathbf{I}_2 = \mathbf{YV}_2 \quad \text{or} \quad \frac{\mathbf{V}_2}{\mathbf{I}_2} = -\frac{\mathbf{V}_1}{\mathbf{I}_1} = \sqrt{\frac{\mathbf{Z}}{\mathbf{Y}}} = \mathbf{Z}_k \tag{10.80}$$

$\mathbf{Z}_k$ has all the properties of a characteristic impedance, for as we proceed down the line it is the constant ratio of $\mathbf{V}$ to $\mathbf{I}$ for each travelling wave and so for a line of infinite length with only one wave generated, say from left to right, it is the input impedance at any point looking away from the generator. If, instead, the line is finite but terminated by a load $\mathbf{Z}_k$, the ratio of $\mathbf{V}$ to $\mathbf{I}$ is still $\mathbf{Z}_k$ even at the load and no reverse wave occurs: the input impedance is thus $\mathbf{Z}_k$ for a correctly terminated line.

Special Cases. At high frequencies (1 Mc/s and above) R and G

* Some prefer to define $\mathbf{P} = \sqrt{-\mathbf{YZ}}$ as a propagation constant so that $\mathbf{P} = j\gamma$.

become negligible compared with ωL and ωC so that $\gamma = j\omega\sqrt{LC}$ and $\mathbf{Z}_k = \sqrt{L/C}$. Thus the attenuation is negligible and because the velocity $1/\sqrt{LC}$ is independent of frequency, dispersion does not occur. High frequency lines are thus effectively loss-less and from the expressions for L and C already derived (equations (5.10), (5.11), (9.22) and (9.23)) waves are transmitted with a velocity $1(\varepsilon_0\mu_0)^{1/2}$ in coaxial cables or twin cables (Lecher lines). This velocity is almost exactly 3×10^8 m/s, the velocity of light *in vacuo*: the effect of dielectric and magnetic media is to lower this to $1/(\varepsilon_r\varepsilon_0\mu_r\mu_0)^{1/2}$, the velocity of light in such media. Note that $\mathbf{Z}_k$ is purely resistive.

At lower frequencies, particularly those used in telegraphy and telephony, the losses cannot be ignored. If, however, it were possible to arrange that

$$R + j\omega L = y^2(G + j\omega C) \tag{10.81}$$

y being any constant, γ would be $y(G + j\omega C)$. The attenuation constant would thus be yG and the velocity $1/yC$, both independent of frequency, and the line would be distortionless. Unfortunately (10.81) can only be achieved if $R = y^2 G$ and $L = y^2 C$, i.e. if $RC = LG$, and in practice $RC \gg LG$. To increase G would introduce unacceptable attenuation and the best which can be done is to increase L either by winding permalloy tape round the conductor (for long low-frequency telegraphy lines) or by inserting coils in series with the cables at intervals of a fraction of a wave-length (for telephony). This does not achieve (10.81) but produces some improvement.

Standing Waves in Loss-less or H-F Lines. When $\gamma = j\beta$, the solution (10.77) becomes

$$\mathbf{I} = \mathbf{I}_1 e^{j\beta x} + \mathbf{I}_2 e^{-j\beta x}; \qquad \mathbf{V} = \mathbf{V}_1 e^{j\beta x} + \mathbf{V}_2 e^{-j\beta x} \tag{10.82}$$

in which $\mathbf{V}_1 = -\mathbf{Z}_k\mathbf{I}_1$ and $\mathbf{V}_2 = \mathbf{Z}_k\mathbf{I}_2$ by (10.80). If, as in Fig. 10.25b, we take the zero to be at the load $\mathbf{Z}_L$ we shall obtain a reflected wave unless $\mathbf{Z}_L = \mathbf{Z}_k$, and from (10.82)

$$\mathbf{Z}_L = \frac{\mathbf{V}_1 + \mathbf{V}_2}{\mathbf{I}_1 + \mathbf{I}_2} \tag{10.83}$$

We define a *reflection coefficient* ρ as $\mathbf{V}_1/\mathbf{V}_2$ at $x = 0$, and it is

clearly also $-\mathbf{I}_1/\mathbf{I}_2$ by (10.80). Hence from (10.83)

$$\mathbf{Z}_L = \frac{\mathbf{V}_2(1+\rho)}{\mathbf{I}_2(1-\rho)} = \mathbf{Z}_k \frac{(1+\rho)}{(1-\rho)} \tag{10.84}$$

or

$$\rho = \frac{\mathbf{Z}_L - \mathbf{Z}_k}{\mathbf{Z}_L + \mathbf{Z}_k} \tag{10.85}$$

The current and potential difference (10.82) are thus given by

$$\mathbf{I} = \mathbf{I}_2(e^{-j\beta x} - \rho e^{j\beta x}); \qquad \mathbf{V} = \mathbf{Z}_k \mathbf{I}_2(e^{-j\beta x} + \rho e^{j\beta x}) \tag{10.86}$$

where $\mathbf{I}_2$ is determined by the generator.

The input impedance at any point a distance l from the load is $\mathbf{V}/\mathbf{I}$ evaluated at $x = -l$ and so

$$\mathbf{Z}_{in} = \mathbf{Z}_k \frac{(e^{j\beta l} + \rho e^{-j\beta l})}{(e^{j\beta l} - \rho e^{-j\beta l})}$$

which, using (10.85) and $e^{\pm j\beta l} = \cos \beta l \pm j \sin \beta l$, gives

$$\mathbf{Z}_{in} = \mathbf{Z}_k \frac{(\mathbf{Z}_L + j\mathbf{Z}_k \tan \beta l)}{(\mathbf{Z}_k + j\mathbf{Z}_L \tan \beta l)} \tag{10.87}$$

For a quarter-wavelength line, $\beta l = \frac{1}{2}\pi$ and

$$\mathbf{Z}_{in} = \mathbf{Z}_k^{\,2}/\mathbf{Z}_L \tag{10.88}$$

a property which enables it to be used to transform one impedance into another. In particular, if open-circuited, $\mathbf{Z}_{in}$ is zero, while if short-circuited $\mathbf{Z}_{in}$ is infinite, so that a short-circuited quarter-wave line acts at its input end as an open circuit and can be used to support lines without affecting their properties. (See Fig. 10.25c.)

The potential difference given by (10.86) represents two travelling waves in opposite directions superposed and is thus a standing wave. The amplitude of the potential difference is

$$|V| = |Z_k||I_2|[(\rho^2+1)+2\rho \cos 2\beta x]^{1/2} \tag{10.89}$$

The ratio of the maximum value of this to its minimum is known as the voltage standing wave ratio (VSWR) and is $(1+\rho)/(1-\rho)$ or $\mathbf{Z}_L/\mathbf{Z}_k$. Measurement of the standing waves on a line will thus yield not only their wave-length but also the value of the load in terms of the characteristic impedance.

10.11 Non-Sinusoidal E.m.f.s and Currents

In considering a variety of A.C. networks in the last few sections we have assumed that the impressed e.m.f.s have been purely sinusoidal, containing only one frequency. The linearity of the components ensured that only this frequency occurred in any output. In practice, the e.m.f. may be periodic but often not sinusoidal. Nevertheless, Fourier's theorem allows us to represent such e.m.f.s as sums of sinusoidal components whose amplitudes and phases depend on the particular waveform but whose frequencies are multiples of the fundamental frequency or harmonics. For any one such component all the previous work will apply, and we can see from equation (10.1) that if $y = u$ is a solution for the component $f_1(t)$ and $y = v$ is a solution for $f_2(t)$ then $y = u + v$ is a solution for $f_1(t) + f_2(t)$: we need therefore only add the effects of the various harmonics.

However, most linear circuits have properties which vary with frequency so that the relative amplitudes of the components and their phases will change as we move from one point of a network to another, and any output waveform will differ from the input giving rise to it (this is distortion, already referred to in section 10.10). Inductances with reactances ωL, for instance, will tend to accentuate the higher harmonics in the potential difference across it, while condensers will do the opposite. Distortion is, of course, to be minimized where accurate reproduction of speech or music is required but it can often be put to use. We shall consider only the simple series RC circuit as an example of what can be done.

Figure 10.26 shows the circuit, whose time constant is $\tau = CR$, with an input in the form of pulses of duration T repeated at regular intervals. The 'square wave' is equivalent to applying suddenly a steady e.m.f. for a time T and then removing it. An output may be taken either across R or across C and since we have seen that the

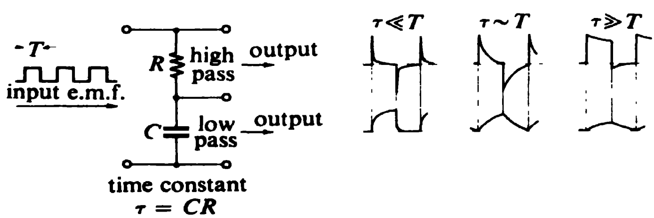


Fig. 10.26. Differentiating and integrating circuits with square-wave input.

latter will be deficient in the higher frequency harmonics it will form a low-pass filter for *any* input, though with no sharp cut-off. For the square wave we apply the results of section 6.5 and obtain the outputs shown in the figure. If $\tau \ll T$, the output from R is a train of very short duration pulses often needed in electronic circuits as triggers, while across C the output almost faithfully reproduces the input waveform. On the other hand if $\tau \gg T$, it is the output from R which tends to reproduce the input.

For a general input potential difference of the form $V(t)$ we know that the sum of $V_R = RI$ and $V_C = Q/C$ is at all times equal to $V(t)$. Moreover, because $I = dQ/dt$, the relation between V_R and V_C is always

$$V_R = \tau \frac{dV_C}{dt} \quad \text{or} \quad V_C = \frac{1}{\tau}\int V_R \, dt \tag{10.90}$$

If τ is very small then for not too rapid changes of potential difference, $V_R \ll V_C$ and the waveform across C reproduces $V(t)$ while V_R is proportional to dV_C/dt which is very nearly $dV(t)/dt$. Conversely, if τ is very large, V_R reproduces $V(t)$ and V_C is proportional to its integral. Figure 10.26 exhibits these properties. A condenser and resistance in series can therefore be used as a simple differentiating or integrating circuit.

10.12 A.C. Measurements

Because there exists no accurate and stable standard source of alternating e.m.f. corresponding to the standard cell, measurements of alternating potential difference and current are less accurate than D.C. measurements and are generally made with meters of various types discussed in section 16.5. In this section we deal with impedance measurement, which falls into three broad classes according to the frequency range. At low frequencies bridge methods are capable of high precision and they can be used with special precautions up to 50 or 100 Mc/s, but at radio-frequencies resonance methods are more common, while in the centimetre-wave region (1,000 Mc/s and above) VSWR methods are usually used. Some of these methods also incidentally provide means of obtaining frequency or wave-length.

Components in Practice. We have already discussed the general construction and properties of resistors, condensers and inductors in previous chapters (sections 6.7, 5.4 and 9.12), but although a

component may be designed as one of these only, it inevitably possesses the properties of all three when used with A.C.

Wirewound resistors possess considerable inductance when simply wound on a bobbin and a small capacitance because of the potential difference between adjacent turns. Non-inductive windings ensure that adjacent wires carry opposing currents as far as possible (as in the bifilar type in which the wire is doubled back on itself) but since this often increases the self-capacitance fairly complex windings must be used. At high frequencies, the skin effect (section 9.6) causes an increase in resistance because of the reduction in the effective cross-section of the wire: this is minimized by using multistranded wire with strands of very small diameter.

Inductors possess an inherent resistance and a self-capacitance for the same reason as resistors and they are thus equivalent to the network of Fig. 10.12. Because of this they are less convenient than condensers as standards in A.C. bridges, particularly when continuous variation is required.

Capacitors or condensers usually have dielectric losses which can be represented by a resistance r in series with the capacitance. The *loss angle* δ is another way of expressing the power factor and is given by

$$\tan \delta = \omega C r \tag{10.91}$$

A.C. Bridges. An ideal bridge as shown in Fig. 10.27a is driven by an alternating source (usually a valve oscillator) and uses a detector which may be a vibration galvanometer (low frequencies up to a few hundred c/s), a telephone receiver (audio-frequencies) or a heterodyne receiver (high frequencies). For complete balance of the bridge the points B and D must be at the same potential *at all times* which means that both amplitude and phase of the alternating potentials at these points must be equalized. In general therefore we expect two balance conditions to emerge.

By the same argument as used for the D.C. Wheatstone bridge, the balance condition is

$$\frac{\mathbf{Z}_1}{\mathbf{Z}_2} = \frac{\mathbf{Z}_3}{\mathbf{Z}_4} \tag{10.92}$$

On cross-multiplying and equating real and imaginary parts two conditions are obtained which are achieved by alternately varying two components in the network. As an example, the Maxwell bridge of Fig. 10.27b is used for determining the inductance L and

resistance r of a coil, and the balance condition is

$$\frac{r+j\omega L}{R} = P(1/Q+j\omega C)$$

giving

$$L = CPR \quad \text{and} \quad r = PR/Q \tag{10.93}$$

These conditions illustrate two properties which are usually desirable in a bridge. First, they are independent of frequency so that

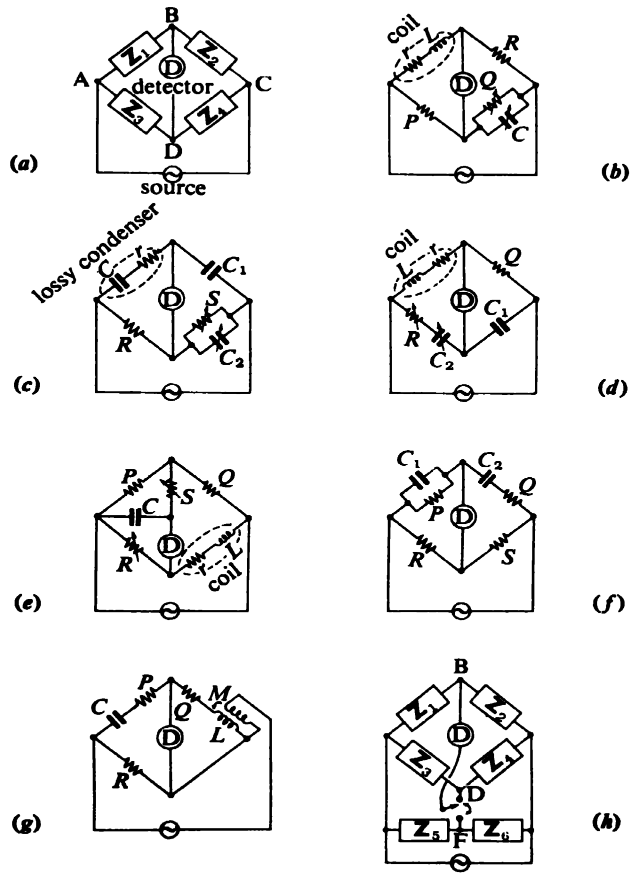


Fig. 10.27. A.C. bridges.

the presence of harmonics in the source does not matter. Secondly, they can be independently satisfied by choosing Q and C as variables since a change in C does not affect the second condition, nor a change in Q the first. In addition, only condensers and resistors are needed as standards.

It can be seen that the solution of a bridge is merely an exercise in the manipulation of complex numbers and we shall only give the circuits and balance conditions for the more important bridges:

Schering bridge (Fig. 10.27c) for capacitance and power factor or for dielectric constant and loss:

$$C = C_1 S/R; \qquad r = C_2 R/C_1$$

and hence
$$\tan\delta = \omega C_2 S$$

Owen's bridge (Fig. 10.27d) for L in terms of C:

$$L = QRC_1; \qquad r = C_1 Q/C_2 \quad \text{(variables } R \text{ and } C_2 r\text{)}$$

Anderson's bridge (Fig. 10.27e) for L in terms of C:

$$L = CR(Q+S+QS/P); \qquad r = QR/P$$

(use the delta-Y transformation to solve it)

Wien's bridge (Fig. 10.27f) for frequency:

$$\omega^2 = 1/QPC_1C_2; \qquad C_1/C_2 = S/R-Q/P$$

Carey-Foster's bridge (Fig. 10.27g) for mutual inductance at low frequencies:

$$M = QRC; \qquad L = M(P+R)/R$$

(use the mutual inductance transformation of (9.43)).

To ensure high sensitivity the detector should be matched to the bridge. A non-zero and incorrect balance will be produced if the source induces e.m.f.s directly in a bridge arm or in the detector through external magnetic fields or through stray capacitance. To avoid this, careful arrangement of the apparatus is necessary and, at high frequencies, shielding of components and leads. Because one side of both source and detector may be earthed and short-circuit an arm, a transformer is often used between one of them and the bridge.

Capacitances to earth from points B and D are difficult to remove and clearly cause a leakage current upsetting the balance. One method of dealing with this is the Wagner earthing device of Fig. 10.27h. $\mathbf{Z}_5$ and $\mathbf{Z}_6$ are additional impedances whose common

connection is earthed. Balance is obtained with both the $\mathbf{Z}_3\mathbf{Z}_4$ pair and with the $\mathbf{Z}_5\mathbf{Z}_6$ pair when the points B, D and F will all be at earth potential, though neither B nor D will be actually connected to earth.

For further details of bridges consult the references given at the end of the chapter.

Resonance Methods. Frequency, inductance and capacitance can be measured very simply by using the properties of the series *LCR* circuit (problem 10.30). Typical of these methods is the *Q*-meter (Fig. 10.28) in which a small resistance R ($\ll r$) is in parallel with a calibrated variable condenser together with a coil whose $\hat{Q}$ and L are required. A voltmeter V with an effectively infinite

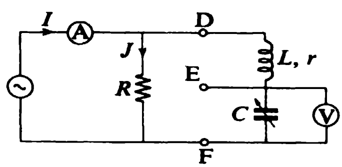


Fig. 10.28. The Q-meter.

resistance is connected across C. C is adjusted so that the meter A indicates resonance and from section 10.6 we know that the voltage across C is $\hat{Q}_0 V$ or $\hat{Q}_0 RI$ because most of the current passes through R, and I and J are nearly equal. For a given instrument the value of I used is adjusted to a set value and the scale of V can be calibrated directly to give $\hat{Q}_0$. In addition, the calibration of C allows L to be calculated at the known frequency. An unknown capacitance can be determined by connecting it across EF and finding the change in C needed to retune to resonance, while its power factor is obtained from the change in $\hat{Q}_0$.

VSWR Methods. Measurement of the standing waves in a mismatched coaxial line enables wave-length and the load impedance to be obtained from relations given in section 10.10. A section of the line has a lengthwise slot cut in the outer conductor and a small probe projecting into the field between the two conductors assumes a potential measured by a rectifier and meter (section 16.5).

Appendix 10.1 Phase–Amplitude ('Vector') Diagrams and Complex Numbers

A common problem encountered in physics is that of adding a number of scalar quantities varying sinusoidally with time at the

same frequency but having various amplitudes and phases. The methods used will be illustrated by adding two such quantities, $A \sin(\omega t+\alpha)$ and $B \sin(\omega t+\beta)$: the extension to more than two is trivial.

Trigonometrical Method. By simple trigonometry

$$A \sin(\omega t+\alpha)+B \sin(\omega t+\beta) = C \sin(\omega t+\gamma)$$

if and

$$\left.\begin{aligned} C &= (A^2+B^2+2AB\cos(\beta-\alpha))^{1/2} \\ \tan\gamma &= \frac{A \sin\alpha+B \sin\beta}{A \cos\alpha+B \cos\beta} \end{aligned}\right\} \qquad (10.94)$$

and it is only C and γ which are needed. The tollowing are two methods of obtaining them used in A.C. theory.

Geometrical Method. Representing the quantities by lines with a length equal to their amplitudes and angles made with a base line equal to their initial phase angles α and β (Fig. 10.29), the resultant C and γ of (10.94) are given by completing the parallelogram or

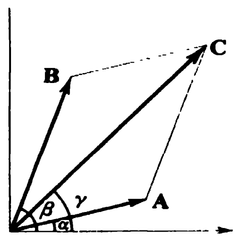


Fig. 10.29. ***Phase–amplitude diagram for the addition of sinusoidally-varying quantities (phasors).***

triangle *as if* A and B were vectors. The diagram is strictly a phase–amplitude diagram as encountered in wave optics although common usage refers to it as a 'vector' diagram. We shall refer to the sinusoidally-varying quantities as *phasors* and the diagram as a *phasor diagram.*

Should instantaneous values be required, they are given by imagining the whole diagram to rotate counterclockwise with an angular velocity ω and finding the projections on a stationary base line.

Complex Number Method. The geometrical method suggests that the addition could be performed by treating the phasor diagram

as an Argand diagram with the imaginary axis vertical if the real axis is horizontally to the right. But, independently of this, if $A\sin(\omega t+\alpha)$ is represented by the complex number $Ae^{j\alpha}$ where $j^2=-1$, and $B\sin(\omega t+\beta)$ by $Be^{j\beta}$, then because $e^{j\alpha}=\cos\alpha+j\sin\alpha$ the resultant of A and B would be given by

$$Ce^{j\gamma} = (A\cos\alpha+B\cos\beta)+j(A\sin\alpha+B\sin\beta) \qquad (10.95)$$

and because the amplitude or modulus of a complex number $a+jb$ is $(a^2+b^2)^{1/2}$ and its argument is $\tan^{-1} b/a$, (10.95) leads straight to (10.94), and we express the addition in the form

$$\mathbf{C} = \mathbf{A}+\mathbf{B} \qquad (10.96)$$

where the bold-face type indicates a complex representation. The addition of phasors can thus be carried out by adding the corresponding complex quantities.

If, as in the geometrical method, instantaneous values are required, then the complex form should be multiplied by $e^{j\omega t}$ (corresponds to rotating the line) and the real or imaginary part taken (corresponds to projection on a base line). The linearity of the differential equations governing A.C. networks ensures that, if $I_0\sin(\omega t+\alpha)$ and $I_0\cos(\omega t+\alpha)$ are both solutions, then so is $I_0e^{j(\omega t+\alpha)}$ and this latter form is often used in the theory of vibrations and waves to represent sinusoidally-varying quantities.

The ratio of two complex numbers is itself complex but, because any time variation represented by $e^{j\omega t}$ must cancel, the result is *not* a phasor. Examples in this chapter are the impedance $\mathbf{Z}$ and admittance $\mathbf{Y}$.

A useful relation to remember is that if $\mathbf{Z}=(a+jb)/(c+jd)$ then

$$|Z| = \left(\frac{a^2+b^2}{c^2+d^2}\right)^{1/2} \quad\text{and}\quad \arg\mathbf{Z} = \tan^{-1}\frac{bc-ad}{ac+bd} \qquad (10.97)$$

References

Mechanical and electric vibrations: Barker (1964). A.C. bridges: Hague (1946). A.C. measurements generally: Harris (1952), Stout (1960).

PROBLEMS

SECTION 10.1

10.1 The current in a non-linear device is given to a sufficient degree of accuracy by $I=a+bV+cV^2$ where V is the potential difference and a, b and c are constants. If the applied potential difference has the form

$V_0 + V_1 \sin \omega t$, show that the effect of the alternating component on the current is to increase the mean value by $\frac{1}{2}cV_1^2$ and to introduce a harmonic of angular frequency 2ω.

10.2 If the applied potential difference across the non-linear device of problem 10.1 is $V_0 + V_1 \sin \omega_1 t + V_2 \sin \omega_2 t$, show that terms involving sum and difference frequencies occur in the current in addition to those expected from the results of the last problem.

SECTION 10.2

10.3 A condenser of capacitance 1 μF is discharged through a coil of self-inductance 10 mH and resistance 200 Ω. At what time after the discharge begins does the current reach its maximum value?

10.4 Condensers of capacitance 0·5 μF and 0·1 μF and an inductor of self-inductance 12 mH and negligible resistance are connected in series. Find the natural angular frequency for discharge of the 0·5 μF condenser by the methods of section 10.2 without assuming that the condensers can be replaced by a single equivalent one.

10.5 Put $CR^2/4L = \delta$ in (10.19) and (10.21) and show how ω_N and Λ vary with δ when damping is small.

SECTION 10.3

*10.6 Show that the natural angular frequencies of an oscillatory discharge of one of the condensers in the network of Fig. 10.30 are ω_0 and $\sqrt{3}\,\omega_0$, where $\omega_0 = 1/(LC)^{1/2}$. (Resistances are to be neglected.)

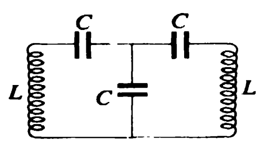


Fig. 10.30. Network for problem 10.6.

SECTION 10.4

10.7 At what angular frequency is the impedance of a capacitance C and a resistance R in series twice that of the same C and R in parallel?

10.8 A 25 W 100 V lamp is to be used on alternating current mains of 200 V RMS and 50 c/s frequency. What is the capacitance of the condenser which, if placed in series with the lamp, will suitably adjust the current flowing, no ohmic resistance being added?

10.9 Find the conductance and susceptance of (*a*) L and R in parallel, (*b*) L and R in series.

10.10 When connected to a D.C. supply, the potential difference across a coil measured by a voltmeter is 1 V when the current through it measured by an ammeter is 1 mA. The same coil across a 50 c/s A.C. supply gives A.C. meter readings of 10 V and 4 mA. What are the resistance and inductance of the coil, assuming that the meters are ideal?

10.11 Find the complex impedance of a coil with self-inductance L and resistance R in parallel with a condenser of capacitance C with an imperfect dielectric equivalent to a series resistance also R. Find (*a*) a value of R which makes the impedance purely resistive at all frequencies and (*b*) a value of ω which makes the impedance purely resistive for all values of R.

10.12 A coil of self-inductance L and resistance R is connected in series with a key and the combination is placed in parallel with a condenser of capacitance C. An alternating source of e.m.f. is applied across C and the current supplied by this source is measured by a milliammeter. The value of C is adjusted until the milliammeter reading remains constant when the key is opened or closed. Show that $2\omega^2LC=1$, using both the algebraic method with complex quantities and a vector diagram. (Turner's method for measuring a large self-inductance.)

10.13 In a conventional A.C. bridge network (Fig. 10.27a), $\mathbf{Z}_1=\mathbf{Z}_2=R$, $\mathbf{Z}_3$ is a condenser C, and $\mathbf{Z}_4$ a variable resistance S. If the detector is removed, draw a vector diagram of potential differences and currents in the network (using the head-to-tail diagram of Fig. 10.9c for $\mathbf{V}$), and show that the phase of the output across BD relative to the input can be varied from 0 to π, without change of amplitude, by varying S. (Phase shift network.)

SECTION 10.5

10.14 An alternating current $I_1 \sin \omega t$ is superimposed upon a direct current I_2. What is the R.M.S. current?

10.15 Two alternating currents of the same frequency but with a phase difference α have peak currents I_1 and I_2 and are fed into a single wire. What is the R.M.S. current?

SECTION 10.6

10.16 A resonant current of 50 mA flows in a circuit consisting of 2 mH, 20 Ω and 0·0003 μF all in series. What is the applied e.m.f.? If the current is reduced to 30 mA by changing the frequency but not the e.m.f., find the new frequency and the phase difference between e.m.f. and current.

10.17 In a series LCR circuit with an applied e.m.f. of constant amplitude and variable frequency, the maximum potential differences across the L, the C and the R occur at angular frequencies ω_L, ω_Q and ω_R respectively. Show that $\omega_Q{}^2=\omega_0{}^2(1-1/2\hat{Q}_0{}^2)$, $\omega_R=\omega_0$ and $\omega_L{}^2=\omega_0{}^2/(1-1/2\hat{Q}_0{}^2)$ where $\omega_0=1/(LC)^{1/2}$.

10.18 Show that the curve of current against angular frequency ω for a series LCR circuit to which a constant alternating e.m.f. is applied is not symmetrical about ω_0, but that the curve of current against $\log_e \omega$ is symmetrical about $\log_e \omega_0$.

10.19 An alternating e.m.f. is applied across the central capacitance of Fig. 10.30. Show that the resonant frequencies of the network are the same as the natural frequencies obtained in problem 10.6.

SECTION 10.7

10.20 Show that the two peaks in the $|I_s|$ against ω curve of Fig. 10.9 are of equal height V/R for ω near ω_0.

10.21 Use the method of section 10.7 to find the condition that $|I_p|$ of (10.67) shall have two maxima and one minimum. Show that this condition is already fulfilled at critical coupling $k = 1/\hat{Q}_0$.

SECTION 10.9

10.22 If $\mathbf{Z}_{oc}$ and $\mathbf{Z}_{sc}$ are the input impedances of a symmetrical T-section when the output terminals are open-circuited and short-circuited respectively, show that $\mathbf{Z}_{oc}\mathbf{Z}_{sc} = \mathbf{Z}_k{}^2$. Verify that the same condition applies to a symmetrical π-section and to a loss-less transmission line (using (10.87)).

10.23 A half T-section ($\frac{1}{2}\mathbf{Z}_1$ in series, $2\mathbf{Z}_2$ in parallel) terminates a T-section ladder network with elements $\mathbf{Z}_1$ and $\mathbf{Z}_2$. Show that the network is correctly terminated if a load with impedance equal to $\mathbf{Z}_{k\pi}$ given by (10.70) is connected across the half-section.

10.24 Show that the cut-off frequency for the high-pass filter of Fig. 10.24b is given by $\omega_0 = 2/(LC)^{1/2}$.

10.25 Show that the pass band of the filter in Fig. 10.24c is from

$$\omega_1 = (\sqrt{2}-1)/(LC)^{1/2} \quad \text{to} \quad \omega_2 = (\sqrt{2}+1)/(LC)^{1/2}$$

10.26 Show that the shape of the attenuation curve (α against ω) for the low-pass filter of Fig. 10.24a is given by $\cosh\alpha = \frac{1}{2}\omega^2 LC - 1$. If a rectifier output with a 100 c/s ripple is to be smoothed, what attenuation in power per T-section is achieved if $L = 20$ H and $C = 5$ μF?

SECTION 10.10

10.27 Assuming that a network like that of Fig. 10.24a behaves as a transmission line at frequencies well below the cut-off, find the time delay per section if $L = 25$ H and $C = 1$ μF. (Artificial delay line.)

10.28 Find the characteristic impedance of a loss-less coaxial cable in which the diameter of the inner conductor is $\frac{1}{3}$ of that of the inner surface of the outer conductor. If the voltage standing wave ratio in the cable is 2·5, what is the resistance of the termination?

SECTION 10.12

10.29 In a series LCR circuit, the alternating e.m.f. is of fixed amplitude and frequency while the capacitance is variable. Show that the curve of current against C has a half-power width δC and a peak value at C_0, where $2C_0/\delta C = \hat{Q}_0$ provided $\hat{Q}_0$ is much greater than 1.

10.30 In an LCR series circuit, the L is the secondary of a transformer feeding in an e.m.f. of fixed amplitude and frequency. The condenser C is variable and calibrated and has a voltmeter of effectively infinite resistance placed across it. The voltmeter reading is a maximum when $C = 5{\cdot}42$ μF. When an unknown capacitance is connected in parallel with C, the voltmeter reading is a maximum for $C = 3{\cdot}15$ μF. What is the unknown capacitance? How would you proceed to measure the unknown if it was larger than the maximum value of C?

CHAPTER 11

THE MOTION OF CHARGED PARTICLES IN ELECTRIC AND MAGNETIC FIELDS

When the pressure in an enclosed space is reduced sufficiently, mean free paths for charged particles, whether ions or electrons, become very large and motion can be controlled by applying electric and magnetic fields. In this chapter we consider general principles of such motion based on the force laws already derived:

$$\mathbf{F} = Q\mathbf{E} \tag{11.1}$$

for an electric field alone (equation (3.7)) and

$$\mathbf{F} = Q\mathbf{v}\times\mathbf{B} \tag{11.2}$$

for a magnetic field alone (equation (8.32)). When both **E** and **B** are present, we shall assume that superposition applies and therefore that

$$\mathbf{F} = Q(\mathbf{E}+\mathbf{v}\times\mathbf{B}) \tag{11.3}$$

often known as the Lorentz force-formula. We also assume that the fields are unaffected by the presence or motion of the charges in them.

For small velocities the mass of a particle may be assumed constant and classical mechanics applied to calculate the paths followed. As the velocities v approach that of light, c, we must use relativistic mechanics according to which the mass m is given by

$$m = \frac{m_0}{(1-v^2/c^2)^{1/2}} \tag{11.4}$$

where m_0 is the *rest mass*. It is outside the scope of this book to justify laws of mechanics but experimental evidence for (11.4) is presented in section 11.3.

11.1 Steady Electric Fields

Because the force on a charged particle in an electric field is given by $\mathbf{F}=Q\mathbf{E}$, any component of velocity perpendicular to **E** is un-

changed while one parallel to **E** increases or decreases according to the sign of Q, i.e. acceleration, of magnitude QE/m, takes place only in the direction of **E**.

Uniform Electric Fields. In a uniform **E**, a charged particle will move exactly as does a projectile in a uniform gravitational field and the path will in general be a parabola.

If the velocity is initially *parallel* to **E** (Fig. 11.1a), acceleration

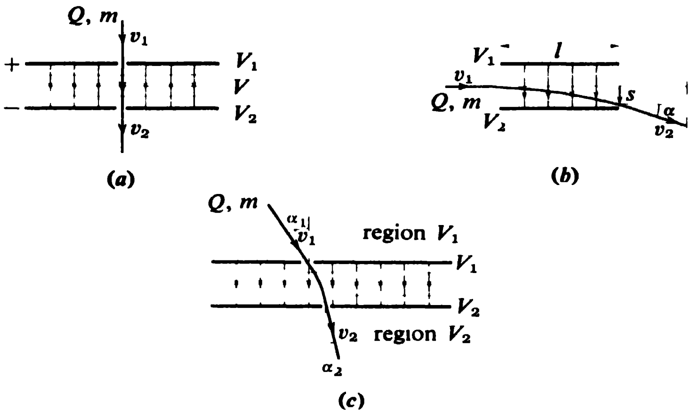


Fig. 11.1. Positive charged particles in an electric field.

occurs without deflection (section 3.7) and the increase in kinetic energy, by (3.25), will be

$$\delta K = \tfrac{1}{2}mv_2{}^2 - \tfrac{1}{2}mv_1{}^2 = Q(V_1 - V_2) \qquad (11.5)$$

This can be used (a) to measure the kinetic energy of charges emitted from a source by retarding them with an opposing field until $v_2 = 0$ when the initial kinetic energy is given by $Q(V_1 - V_2)$, (b) to produce a beam of charges of approximately known and uniform energy from a source emitting them with negligible velocity: in this case $v_1 = 0$ and the final energy is $Q(V_2 - V_1)$. The production of high energy particles using (b) is further discussed in section 11.4.

When the initial velocity v_1 is *perpendicular* to **E** (Fig. 11.1b) a particle spends a time l/v_1 undergoing a transverse acceleration QE/m. The final velocity is thus the resultant of v_1 perpendicular to **E** and QEl/mv_1 parallel to **E** so that

$$\tan\alpha = QEl/mv_1{}^2 = Q\frac{l}{d}\frac{V}{2K} \qquad (11.6)$$

where K is the initial kinetic energy and the field is produced by a potential difference V between plates a distance d apart. The linear deflection in the field is given by

$$s = \frac{1}{2}\frac{QE}{m}\frac{l^2}{v_1^2} = Q\frac{l^2}{d}\frac{V}{4K} \tag{11.7}$$

This is used (*a*) to deflect a beam of electrons in cathode-ray tubes by an amount proportional to V, an externally applied potential difference, (*b*) to determine the ratio e/m for cathode rays (Thomson, 1897) and thermionically emitted charges (thermoelectrons, Thomson, 1899) although this requires a beam of known and uniform v_1, and (*c*) to separate from a beam of particles with the same charge but different energies those with a particular K.

In the most general case (Fig. 11.1c) the particle enters the field from an equipotential region V_1 with velocity v_1 and leaves with velocity v_2 in a region with potential V_2. The following apply:

$$\tfrac{1}{2}mv_1^2 + QV_1 = \tfrac{1}{2}mv_2^2 + QV_2 \quad \text{from (3.25)} \tag{11.8}$$

$$v_1 \sin\alpha_1 = v_2 \sin\alpha_2 \tag{11.9}$$

If the charges are originally produced by acceleration from rest at a point where the potential is taken as zero then both sides of (11.8) are zero and hence

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{v_2}{v_1} = \sqrt{\frac{V_2}{V_1}} \tag{11.10}$$

Non-uniform Electric Fields. If the thickness of the uniform field in Fig. 11.1c is very small, the path of the particle is clearly refracted in the same manner as a ray of light at the interface between two media. Equation (11.10) shows that the quantity corresponding to refractive index is v or $V^{1/2}$. Non-uniform fields are analogous to optical media with continuously variable refractive indices and the surfaces labelled V_1 and V_2 in the figure can be considered as two typical equipotentials whose separation is small enough for the field between them to be considered approximately uniform. Equation (11.10) thus gives the relation between the direction of the particle (defined by α), its velocity and the potential for motion in any electric field.

Beams of charged particles can therefore be deviated in ways analogous to the refraction of light: when the trajectories are those of electrons we have a branch of *electron optics*. Note, however,

that whereas positive charges increase their velocity as V decreases, electrons will suffer an increase in v as V increases (in contrast to light). Figure 11.2a illustrates the difference in behaviour of positive and negative charges.

The focusing properties of fields, which we shall encounter several times, are important for two reasons. The first is that

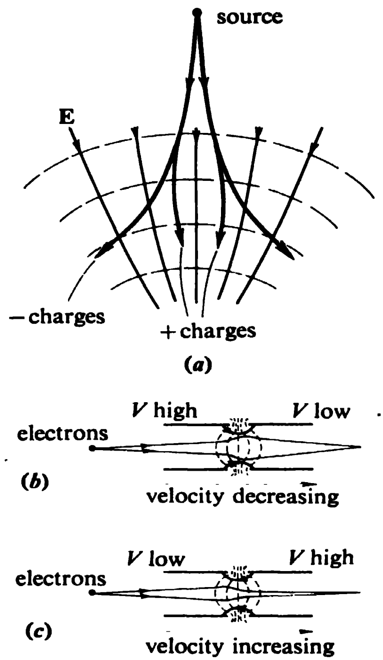


Fig. 11.2. Charged particles in non-uniform electric fields. The dotted lines are equipotentials.

beams of charge of a single sign are naturally defocusing because the electrostatic repulsion exceeds the magnetic attraction (problem 11.1), and the second is that a slit used to define a beam has a finite width so that even with what is effectively a point source, the beam is naturally divergent. Focusing with electrostatic lenses is common in cathode-ray tubes and a typical form is shown in Figs. 11.2b and c. The two cylindrical electrodes produce equipotential surfaces and lines of force as shown and, even though one half is divergent, the whole is convergent because the electrons spend

more time in that part of the field with the lower potential. In a cathode-ray tube the focusing is controlled by altering the relative potentials of the two cylinders.

Cylindrical Electric Fields. The electric field between two coaxial cylindrical conductors is inversely proportional to the distance from the axis (Fig. 11.3 and equation (4.14)). If a beam of particles enters such a field perpendicular to **E** and a distance r from the axis, and if in addition v satisfies the condition

$$\frac{mv^2}{r} = QE \tag{11.11}$$

then the radius of the path traversed will also be r, the velocity will remain perpendicular to **E** and (11.11) will continue to apply. If

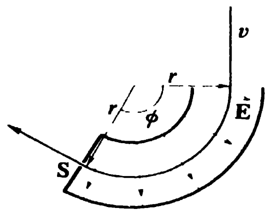


Fig. 11.3. Path of an electron in a cylindrical electric field.

$E = A/r$ where A depends on the potential difference and radii of the cylinders, $mv^2 = QA$ or $K = 2QA$ and only those particles with this kinetic energy will emerge through the slit S. It can be shown that if the beam entering is slightly divergent the field will focus the trajectories at S for an angle $\phi = \pi/\sqrt{2}$ or about 127°.

11.2 Steady Magnetic Fields

In contrast to the electric field, equation (11.2) ($\mathbf{F} = Q\mathbf{v} \times \mathbf{B}$) shows that the velocity component parallel to **B** is unchanged, acceleration being always at right angles both to velocity and field.

Uniform Magnetic Fields—Circular Paths. Consider first the motion of a charged particle whose velocity at some instant is *perpendicular* to a uniform **B** (Fig. 11.4). The force QvB is at right angles to **v** and can thus only alter the direction of motion and not the speed. If the radius of curvature of the path at this instant is r

then because the acceleration is v^2/r

$$QvB = mv^2/r$$

or

$$r = mv/QB = p/QB \tag{11.12}$$

where p is the linear momentum. Because v is constant in magnitude, r is as well, and the path is thus a circle of radius given by (11.12). Proportionality between r and p for a given Q allows particles of various momenta to be distinguished in, say, a cloud

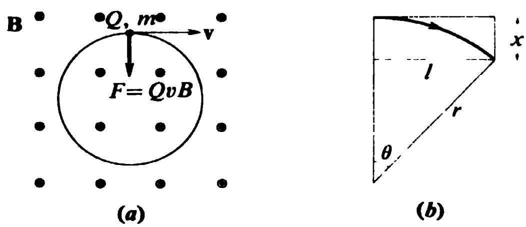


Fig. 11.4. Motion of a positive charged particle projected with a velocity perpendicular to a uniform **B**.

chamber, while if a beam of uniform velocities can be obtained, the proportionality between r and m can be used to compare masses of, or to separate, isotopes.

From Fig. 11.4b, the lateral deflection x after travelling a distance l in the original direction is $r(1-\cos\theta)$ where $\sin\theta = l/r$. For *small* deflections this is

$$x \doteqdot l^2/2r = l^2QB/2mv \tag{11.13}$$

The deflection of an electron beam by a magnetic field produced by current-carrying coils is used in cathode-ray tubes and differs from the electrostatic method in enabling the deflecting system to be outside the tube and therefore accessible. Lenard (1900) used magnetic deflection of photoelectrically emitted particles of known energy to show that they were electrons.

A geometrical property of circular paths is shown in Fig. 11.5a. OA is the diameter ($=2r$) of one circle and OB that of a second, the angle between them being α. P is the intersection of OA with the second circle and it is easily seen that BP is at right angles to OA. The distance AP is $2r(1-\cos\alpha)$ or $4r\sin^2\frac{1}{2}\alpha$. A third circle shown dotted also has a diameter making α with OA which it cuts at P.

If, as in Fig. 11.5b, O is a source of particles all with the same momenta in a uniform magnetic field but with a divergence angle 2α which is small, the extreme trajectories come to a focus at P while the mean trajectory passes through A, the spread AP being $r\alpha^2$, a quantity of the second order of smallness. First-order focusing is thus achieved for 180° paths in a way rather similar to that for the 127° trajectories in a cylindrical electric field.

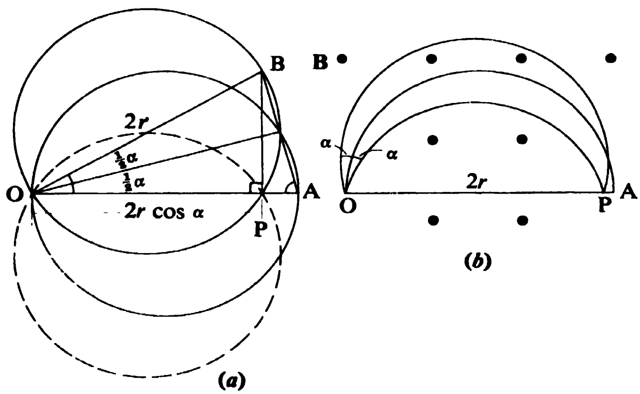


Fig. 11.5. First-order magnetic focusing.

Probably the most important property of (11.12) is revealed by using the angular frequency $\omega = v/r$ so that

$$\omega_c = QB/m \tag{11.14}$$

and the true frequency $QB/2\pi m$ is known as the *cyclotron frequency*, which depends only on the Q/m of the particles and the magnetic flux density and is *independent of radius and velocity* (as long as $v \ll c$).

Uniform Magnetic Fields—Helical Paths. A particle projected *parallel* to **B** experiences no force and thus moves in a straight line with uniform speed.

A particle initially possessing a velocity **v** making an angle α with **B** has velocity components parallel and perpendicular to **B** ($v_{\parallel} = v \cos \alpha$, $v_{\perp} = v \sin \alpha$). The combination of circular motion due to $v_{\perp}$ and the uniform motion due to $v_{\parallel}$ will produce a helix (Fig. 11.6) whose pitch is $(2\pi m v \cos \alpha)/QB$ or $(2\pi p \cos \alpha)/QB$. If a beam consisting of identical particles with the same linear momenta p but

with a small divergence angle α is subjected to a uniform magnetic field along the mean direction of motion, the pitch of all the helices is very nearly $2\pi p/QB$ and again a first-order focus is obtained.

Non-uniform Magnetic Fields. The focusing property of a uniform magnetic field just mentioned is also possessed by the field along the axis of a circular coil and this type of electron lens is used in electron microscopes and in some types of beta-ray spectrograph (Fig. 11.6b).

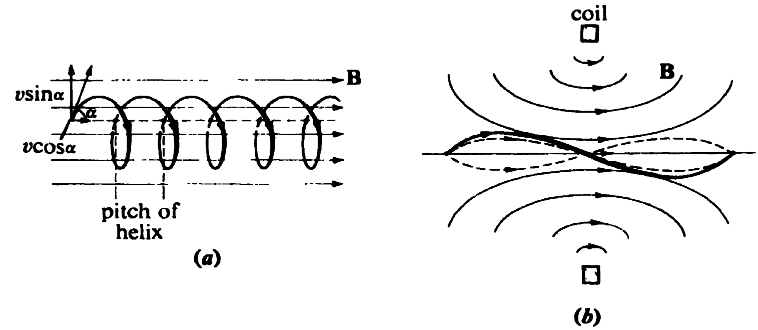


Fig. 11.6. (a) Helical path of an electron in a uniform **B**; *(b) first-order magnetic focusing by a current-carrying coil.*

The particular case of a divergent or convergent magnetic field with cylindrical symmetry is illustrated for *electron* motion in Fig. 11.7. The field **B** has two components: B_z, responsible for the circular orbit, and B_r, producing a force on the electron away from the origin of the lines of force at O and making the electron follow a helical path.

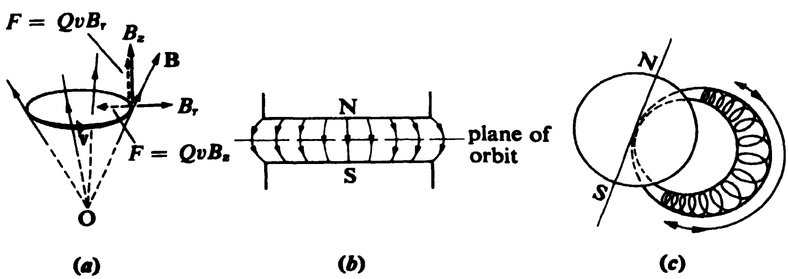


Fig. 11.7. (a) Electron in a non-uniform **B**; *(b) type of non-uniformity retaining particles in plane circular orbits; (c) trapping of particles in a van Allen belt.*

Two applications of this may be mentioned. First, a beam of particles moving in a circular path in a strictly uniform field will lose any particles which move out of the plane of the orbit, and so a non-uniformity of the type shown in Fig. 11.7b is necessary to refocus the beam. The second application concerns the motion of particles in magnetic fields with the configuration of the earth's field: approximately dipolar and converging towards the N and S poles. A charge entering the field will begin to describe a helical path and the non-uniformity is such as to carry it away from the nearest pole. Eventually it reaches a converging field and the pitch of the helix begins to shorten until it becomes zero and then reverses. The particle thus oscillates continuously and is trapped in the field. In the regions of space near the earth a large number of particles are trapped in this way forming the van Allen belts (Fig. 11.7c). The reversal of path in a converging field is known as the *mirror effect.*

11.3 Steady Electric and Magnetic Fields

Parallel Uniform **E** *and* **B**. A charge Q projected with velocity **v** as in Fig. 11.8a experiences the forces shown. In the xz-plane the only force acting is that due to **B** and this will produce a circular

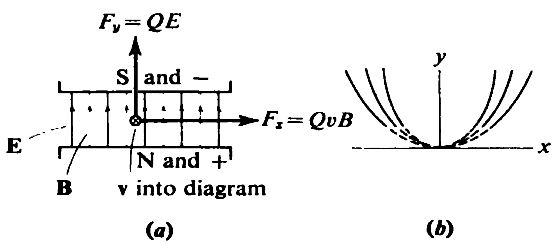


Fig. 11.8. Positive charges in parallel **E** *and* **B**.

trajectory as usual, the x deflection after a distance l along the z-axis being given by (11.13) as $l^2QB/2mv$. In the yz-plane the only force acting is QE and the deflection here is $y = l^2QE/2mv^2$ from (11.7). Eliminating v gives

$$y = \frac{2E}{l^2B}\frac{m}{Q}x^2 \tag{11.15}$$

which means that in a plane of given l, particles of any velocity but

with a definite Q/m will be distributed along a parabola (Fig. 11.8b). This was the method used by Thomson to determine the masses of the positive ions produced in a gaseous discharge. The method has been superseded by modern mass spectrometers (see end of this section) in which beams of given Q/m are not spread out over a parabola with resultant loss of intensity, but are focused.

Perpendicular Uniform **E** *and* **B**. Most applications using simultaneously applied electric and magnetic fields arrange for them to be perpendicular. Using the notation of Fig. 11.9a in which the components of velocity are v_x, v_y and v_z, it can be seen

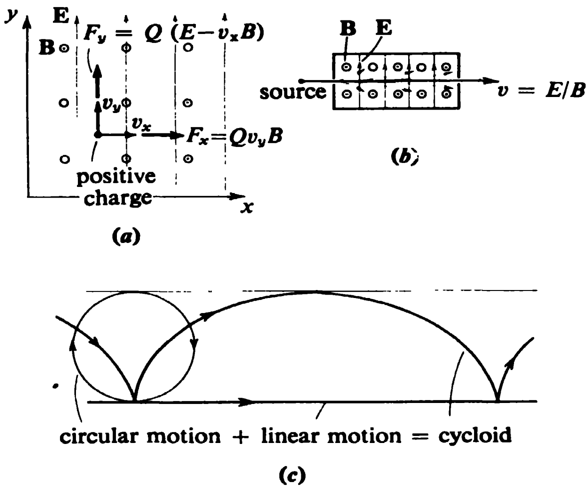


Fig. 11.9. Motion of a positive charged particle in perpendicular uniform **E** *and* **B**.

that any v_z will remain unchanged because no forces act along the z-axis, and only motion in the xy-plane can occur.

An important special case arises when initially $v_y = 0$ and $v_x = E/B$ for then both F_x and F_y are, and remain, zero and the velocity is unchanged. This is the basis of the *velocity selector* or *filter* illustrated in Fig. 11.9b for, of all particles entering crossed fields with a velocity perpendicular to both, only those with $v = E/B$ are undeflected.

The general trajectory for any initial velocity can be obtained by solving the equations of motion $F_y = m\ddot{y} = Q(E - v_xB)$, $F_x = m\ddot{x} = Qv_yB$, but another method is more instructive. It is apparent from

the special case above that a charge travelling with $v_x = E/B$ experiences no force in the y-direction and hence is effectively in a zero y-field. A non-zero y-field will only come into play when v_x differs from E/B, so let us put $v_x = u_x + E/B$ when the equations of motion become $F_y = Qu_xB$, $F_x = Qv_yB$. It is apparent that if we viewed the whole system from an origin moving with an x-velocity of E/B, these would be the equations of motion: that is, the path would be a circle of radius given by (11.12) with v having components u_x and v_y. The resultant motion from a *fixed* origin is thus a combination of uniform circular motion and a uniform linear velocity in the same plane—a cycloid (Fig. 11.9c). The exact type of cycloid depends on the initial velocity.

Bucherer's Experiment. The first convincing direct verification of the variation of mass with velocity was given by Bucherer in 1909. He used a source of β-particles situated at the centre of a pair of circular parallel plates between which electric and magnetic fields were maintained in the directions shown in Fig. 11.10. In any

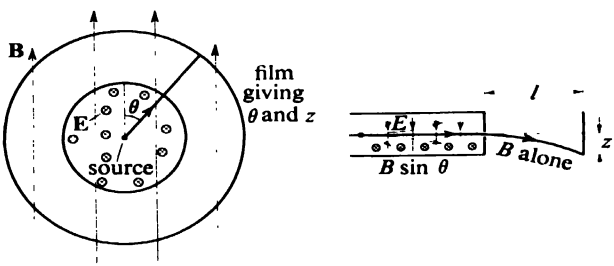


Fig. 11.10. Bucherer's experiment.

direction θ, only those particles for which $QvB \sin \theta = QE$, or $v = E/B \sin \theta$, are undeflected so that the value of θ is a measure of v. On emerging from the plates the particles are subjected only to the magnetic field and although the paths are strictly helical, the displacement z for small deflections is given by (11.13) so that

$$z = (l^2QB^2 \sin \theta)/2mE$$

and this enables Q/m for the particles to be obtained for a range of velocities. Table 11.1 gives Bucherer's results from which it is clear that the relativistic mass given by (11.4) must be used when v/c is not negligible compared with 1.

Table 11.1

$\beta = v/c$	0·317	0·379	0·428	0·515	0·687
e/m(C/kg)	$1{\cdot}66 \times 10^{11}$	1·63	1·59	1·51	1·28
$e/m(1-\beta^2)^{1/2}$	$1{\cdot}75 \times 10^{11}$	1·76	1·76	1·76	1·77

Mass Spectrometry. A mass spectrometer must incorporate a means of producing a beam of particles of uniform velocity and a means of deflecting the beam by an amount dependent on Q/m and focusing it. Uniform velocities can be roughly achieved for low energy sources by accelerating them through a known potential, but some form of velocity selector is better and gives higher resolving power. Subsequent deflection and focusing of the beam often uses a magnetic field occupying only a sector instead of the 180° of Fig. 11.5b: this enables both source and detector to be located outside the field.

11.4 Time-Varying Fields and the Acceleration of Particles

Only electric fields can accelerate charged particles, and the simplest machines for doing so are those which apply an electrostatic field continuously to a particle over its whole trajectory—the Van de Graaff (section 5.9) and Cockcroft–Walton generators are examples. These require the production in the machine of a potential difference in volts equal to the final energy of the particle in eV and thus mean the development of very high potentials.

There are two ways of overcoming this. One is to use the electric field accompanying a changing **B** (in the betatron) and the second is to use an *alternating* field of relatively low potential and shield the particles from its effect during those parts of the cycle when less than maximum acceleration would take place (linear accelerator and cyclotron): both need time-varying fields. At high energies problems connected with the relativistic variation of mass arise and these demand additional refinements.

The Betatron. We have seen in section 9.2 that a changing magnetic flux density is equivalent to an e.m.f. round a closed path and hence produces a non-zero line integral of **E** round a closed path. Consider a magnetic field as in Fig. 11.11 which varies with R but has the same magnitude at all points the same distance from the centre O. Let the flux density at R be B_0. If the magnetic field does not change in time, the momentum of a particle following the circular path is, by $QvB_0 = mv^2/R$,

$$p = QRB_0 \tag{11.16}$$

If the total magnetic flux linking the orbit is Φ, the mean flux density over it is $\bar{B}=\Phi/\pi R^2$. If now Φ changes, an electric field E tangential to the orbit is produced where

$$\text{E.m.f., } E\times 2\pi R = \mathrm{d}\Phi/\mathrm{d}t = \pi R^2\,\mathrm{d}\bar{B}/\mathrm{d}t \quad \text{or} \quad E = \tfrac{1}{2}R\,\mathrm{d}\bar{B}/\mathrm{d}t$$

This would produce a tangential acceleration causing p to change:

$$\mathrm{d}p/\mathrm{d}t = QE = \tfrac{1}{2}QR\,\mathrm{d}\bar{B}/\mathrm{d}t$$

and this would normally alter R by (11.16) were B_0 constant. If, however, B_0 is also changed so that the ratio p/B_0 is preserved then

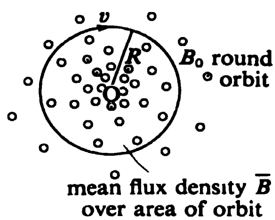


Fig. 11.11. Orbit of charged particle in the betatron.

the radius given by (11.16) remains constant. Thus the condition for a stable orbit is

$$\frac{\mathrm{d}B_0}{\mathrm{d}t} = \frac{1}{QR}\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{1}{2}\frac{\mathrm{d}\bar{B}}{\mathrm{d}t} \tag{11.17}$$

or the rate of change of B *at* the orbit must be half that of the average B *over* the orbit: a field which falls off with increasing radius. The above analysis is still valid at relativistic velocities because it is expressed in terms of p and not of m.

The Linear Accelerator and Cyclotron. The basic form of both these machines enables a high frequency alternating field to accelerate particles for part of the cycle only, while shielding them from the field during the rest of it. In the linear accelerator (Fig. 11.12a) the shielding is performed by electric-field-free cylinders whose length must increase as $n^{1/2}$ as the velocity increases, to preserve the phase relation between the applied R.F. and the presence of the particle in the gaps. In the cyclotron, the electric-field-free regions are those inside two dees (Fig. 11.12b) across which a magnetic field is applied. Because the time spent in a dee is independent of the radius of the path and the velocity (equation (11.14)), a fixed

frequency alternating E of $\omega = QB/m$ will stay in phase with particles. In both applications the times at which the particles cross the gaps are shown in relation to the applied R.F. in Fig. 11.12c.

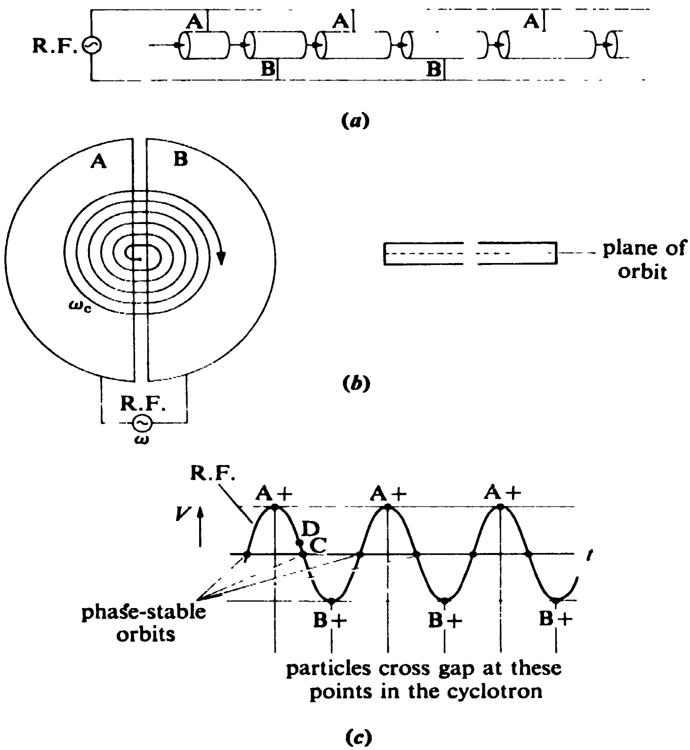


Fig. 11.12. (a) Linear accelerator; (b) cyclotron; (c) relation between R.F. and orbital frequency of particle.

Relativistic Effects. The relativistic increase in mass of a particle, $m - m_0$, can be shown to be equal to the kinetic energy K divided by c^2 or

$$K = (m - m_0)c^2 \tag{11.18}$$

which for small velocities gives $\frac{1}{2}m_0 v^2$. From (11.4)

$$(1-\beta^2)^{1/2} = m_0/m = \frac{1}{1 + K/m_0 c^2} \tag{11.19}$$

showing that the relativistic increase in mass begins to become serious when K/m_0c^2 is not negligible compared with 1. For an electron, $m_0c^2 = 0{\cdot}51$ MeV while for a proton $m_0c^2 = 938$ MeV. This means that the cyclotron frequency

$$\nu_c = \frac{QB}{2\pi m_0(1 + K/m_0c^2)} \qquad (11.20)$$

decreases markedly for quite low energy electrons, and in a fixed-frequency cyclotron they fall out of phase with the applied R.F. until they are crossing the gaps at the times labelled C in Fig. 11.12c. No further increase in energy is possible.

The orbits in which the particles cross the gaps at C are stable, however, in that any further falling out of phase is corrected by acceleration or retardation. In the *synchrocyclotron* the energy in the phase-stable orbit is increased by slowly lowering the frequency of the applied R.F. causing the particles to occupy points in Fig. 11.12c at D relative to the new frequency: this accelerates them until they once more occupy points C but at lower ν_c and higher K (by (11.20)). In the electron *synchrotron* the same effect is obtained by slowly increasing the magnetic flux density while keeping the applied R.F. constant, the energy being high enough for $\beta \doteq 1$ and the velocity thus nearly constant. In the proton synchrotron the particles are still (at 1 GeV) well below the velocity of light and the increasing **B** has to be accompanied by a decrease of frequency because of the increasing velocity. Livingston (1954) gives an excellent account of particle accelerators.

11.5 Magnetic Dipoles in Magnetic Fields

It is now known that the magnetic dipole moments of atoms and of their constituent fundamental particles are always associated with angular momentum. We shall discuss some direct evidence for this in section 14.11 and we see below how even a crude model of the nuclear atom leads to it. For any system the ratio of magnetic moment **m** to angular momentum **L** is known as the *gyromagnetic ratio*, or, more correctly but less euphoniously, the magnetogyric ratio, γ. Thus

$$\mathbf{m} = \gamma \mathbf{L} \qquad \text{(Definition of } \gamma) \quad (11.21)$$

It is the invariable association of **L** with **m** which gives the magnetic dipole quite different characteristics from those of the electric dipole. One difference results from quantum theory: the

angular momentum of any system can only take on certain discrete values, usually expressed in terms of a fundamental unit $h/2\pi$ where h is Planck's constant ($6{\cdot}625 \times 10^{-34}$ J-s): it follows that a similar discreteness will show itself in **m**. The second major difference is that the application of a magnetic field to a magnetic dipole causes it to precess whereas an electric dipole merely tends to align itself along an electric field.

Precession in a Magnetic Field. The reason for precession and a value for its frequency can be obtained in the same way as for a symmetrical spinning top (Fig. 11.13). The weight and normal reaction form a couple **T** acting into the plane of the diagram

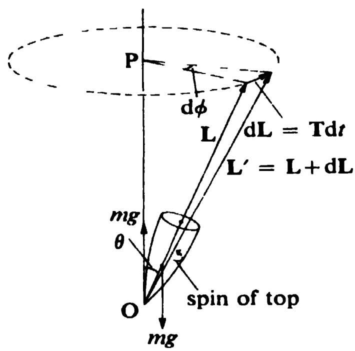


Fig. 11.13. Precession of a spinning top.

(appendix 7.1). In a time dt, **T** causes a change $\mathbf{dL} = \mathbf{T}\,\mathrm{d}t$ in angular momentum just as a force **F** would produce a change **F** dt in linear momentum. The change is also into the plane of the diagram and the addition of **dL** to **L** causes the latter to move into a new position **L**′. The new plane OPL′ is now the one to which the next increment **dL** is normal, and **L** thus traces out a cone and the top precesses.

The angle $\mathrm{d}\phi = \mathrm{d}L/L \sin\theta$ and the angular frequency of precession is therefore

$$\omega = \mathrm{d}\phi/\mathrm{d}t = \frac{\mathrm{d}L}{\mathrm{d}t}\frac{1}{L\sin\theta}$$

$$= \frac{T}{L\sin\theta}$$

If the system is a magnetic dipole of moment $\mathbf{m}$ in a magnetic field of flux density $\mathbf{B}$ along OP then $T_\theta = mB \sin\theta$ and the angular frequency of precession is, taking account of the directions,

$$\omega_L = -mB/L = -\gamma B \tag{11.22}$$

The frequency $\nu_L = \omega_L/2\pi$ is often known as the *Larmor frequency*, a name originally restricted to the precession of orbital electrons.

Magnetic Moment of an Orbital Electron. An electron of charge e and mass m in a plane orbit about a nucleus N (Fig. 11.14) has a constant angular momentum $L = mr^2\omega$, although both r and ω may

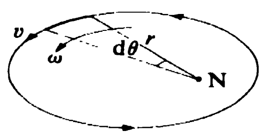


Fig. 11.14. The magnetic moment of an orbital electron.

vary from one part of the orbit to another. The magnetic dipole moment of the electron arises because it is equivalent to a current-carrying loop. If the time for one revolution is τ the electron crosses any point in the orbit $1/\tau$ times per second and is thus equivalent to a mean current $I = e/\tau$. The area S of the orbit is the sum of elementary areas $\frac{1}{2}r^2 \,\mathrm{d}\theta$ so that

$$\mathbf{m} = I\mathbf{S} = \frac{e\hat{\mathbf{n}}}{2\tau}\oint r^2 \,\mathrm{d}\theta$$

$$= \frac{e\hat{\mathbf{n}}}{2\tau}\oint r^2\omega \,\mathrm{d}t$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the plane. Since $r^2\omega = L/m$, a constant,

$$\mathbf{m} = e\mathbf{L}/2m \tag{11.23}$$

so that for an orbital electron the gyromagnetic ratio γ is $e/2m$. Since L is expressed in units of $h/2\pi$, a natural unit for atomic magnetic moments, from (11.23), is $eh/4\pi m$ known as the *Bohr magneton* and denoted by μ_B.

There is evidence (see section 14.11) that the electron has an intrinsic moment, known as its spin moment, for which γ is almost exactly e/m, and an atom as a whole will have a resultant magnetic

moment due to both orbital and spin contributions. We define a quantity known as the Lande g-factor by

$$\gamma = ge/2m \quad \text{(Definition of } g) \quad (11.24)$$

so that $g = 1$ for the magnetic moment due to the orbital motion of a single electron and $g = 2$ for its spin magnetic moment. In a complete atom we therefore expect g to lie between 1 and 2.

Nucleons (protons and neutrons) also have intrinsic magnetic moments and here the *nuclear magneton* $eh/4\pi M$ and the *nuclear g-factor* $(2M\gamma/e)$ are similarly defined but use the proton mass M in place of the electron mass.

11.6 Resonances

Two natural frequencies have been encountered in this chapter: the cyclotron frequency given by (11.14) and the generalized Larmor frequency given by (11.22). Both occur in a steady magnetic field and both lead to important resonance phenomena.

Cyclotron Resonance. A charged particle moving in a circular or helical orbit in a steady magnetic field revolves with an angular frequency $\omega_c = eB/m$ so that a determination of ω_c and B yields the specific charge e/m. When an electric field of angular frequency ω is applied perpendicular to **B** and either ω or ω_c varied, resonance will occur when $\omega = \omega_c$ for then the orbit will expand and the particle will extract energy from the oscillator producing the electric field. In the cyclotron itself, ω and ω_c are fixed and equal and the orbits increase in radius by discrete amounts every half-cycle, but if the particles are free and not shielded by dees the orbits expand continuously (the omegatron—Fig. 11.15).

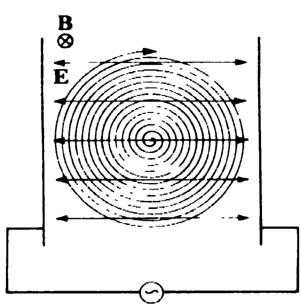


Fig. 11.15. The omegatron.

Sommer, Thomas and Hipple (1951) detected the expansion of the orbits at resonance by means of an ion collector at the edge (proton e/M); while for electrons, Gardner (1951) used helical paths along $\mathbf{B}$ and collected the electron current through a narrow slit—at resonance the expansion of the orbits caused a fall in the current. The magnetic flux density $\mathbf{B}$ was determined in each case by proton magnetic resonance as explained below.

Cyclotron resonance can also be used to find e/m for the carriers of current in metals and semiconductors (section 12.5) provided the mean free paths are long enough (Kip, 1960).

Paramagnetic Resonance. A magnetic dipole precesses in a steady magnetic field with an angular frequency $\omega_L = \gamma B = geB/2m$ so that a determination of ω_L and B yields the gyromagnetic ratio. A small alternating magnetic flux density $\mathbf{B}'$ of angular frequency ω applied perpendicular to $\mathbf{B}$ can be considered as two rotating fields with opposite senses (Figs. 11.16a and b). When $\omega = \omega_L$, one of the

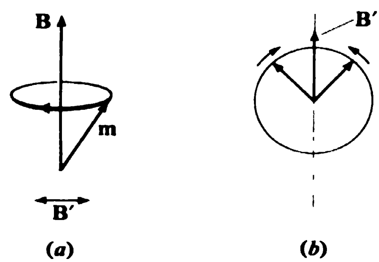


Fig. 11.16. Paramagnetic resonance.

components is rotating in the same sense and with the same frequency as the precessing dipole. If we look at the whole system from a set of co-ordinates rotating with an angular velocity ω_L, both the moment $\mathbf{m}$ and the in-phase component of $\mathbf{B}'$ will appear stationary and $\mathbf{m}$ will thus now precess about $\mathbf{B}'$, changing its angle with $\mathbf{B}$. In doing so it will absorb energy from the oscillating source since the potential energy is $-mB \cos \theta$. In practice ω is often kept constant and B is swept over a small range, thus varying ω_L.

Electron spin resonance (E.S.R.) occurs in substances containing electrons with unpaired spins and provides valuable information about local fields in solids. Nuclear magnetic resonance (N.M.R.)

also gives information about the structure of materials: the gyromagnetic ratio of the proton was determined by Thomas, Driscoll and Hipple (1920) with great precision in terms of a magnetic field measured by the balance of Fig. 8.14b. This enabled γ_p to be measured absolutely and to provide a rapid and accurate method of finding the value of a magnetic flux density using proton resonance.

11.7 Summary of Chapter 11

While no new concepts or laws have arisen in this chapter, the results established are important not only because of their numerous applications in present-day physics but also because the measurements made confirm that we may extend the basic macroscopic force laws to elementary particles.

PROBLEMS

SECTION 11.1

11.1 A beam of electrons from a filament may be regarded as a very long cylindrical region of space of radius a filled with electric charge of uniform density ρ moving with a velocity v. Using the electric and magnetic fields calculated in previous chapters, find the resultant force on any small portion dQ of the moving charge. Show that, if $\varepsilon_0\mu_0 = 1/c^2$ where c is the velocity of light, the force must always be away from the axis. What is the effect if the current consists of equal densities of positive and negative charges?

11.2 The electron beam in a cathode-ray tube is deflected by plates 2 cm long and 5 mm apart. Estimate the linear deflection of the beam at a screen 20 cm away from the plates when 200 V is applied across them. The anode potential of the tube is 500 V.

SECTION 11.2

11.3 What is the radius of the track of a 2 MeV proton in a magnetic field of flux density 1 Wb/m^2?

11.4 Calculate the cyclotron frequency for electrons in a magnetic flux density of 0·2 Wb/m^2.

SECTION 11.3

11.5 A stream of charged particles with various velocities is projected in a direction at right angles both to an electric field of 4 V/cm and a magnetic flux density of 10^{-3} Wb/m^2. What is the speed of the undeflected particles? What energy have they if they are protons?

11.6 Photoelectrons are liberated from a plate by ultra-violet radiation, their initial velocity being negligible. A magnetic flux density **B** is maintained parallel to the plate and an electric field perpendicular to it. The

electric field is produced by a second plate parallel to the first, a distance d from it and at a positive potential V with respect to it. Show that the value of d for which current just fails to pass between the plates is $(2mV/eB^2)$, where e and m are the charge and mass of the electron.

SECTION 11.4

11.7 Calculate β $(=v/c)$ for 100 MeV electrons and 100 MeV protons.

11.8 Prove that the distance between the centres of the gaps in the non-relativistic linear accelerator should increase as $n^{1/2}$.

*11.9 Show that phase stability occurs in the linear accelerator when relativistic velocities are reached if particles cross the gaps while the R.F. voltage is increasing.

SECTION 11.6

11.10 Calculate the Larmor frequencies for a single orbital electron and for a single spinning electron in a magnetic flux density of 0·2 Wb/m^2.

CHAPTER 12

CONDUCTION

The important electromagnetic properties of materials are conductivity, relative permittivity and relative permeability. This chapter and the next two will deal respectively with these, mainly from the macroscopic viewpoint. In each case we shall end by showing briefly how microscopic theory explains the properties in terms of simple atomic models, but because quantum theory is essential if we are to proceed very far we can do no more than indicate the general lines of approach. In this chapter the discussion of conductivity begun in chapter 6 is rounded off.

In one sense any system with two terminals between which a current flows when a potential difference is applied is a conductor but here we shall only consider systems consisting of homogeneous materials. There is no space to deal with devices such as rectifiers, thermionic valves, transistors, some of whose characteristics were shown in Fig. 6.2: many texts on physical electronics give ample details (see references). We first develop some general concepts needed in theories of conduction (mobility, coefficient of diffusion) and then look briefly at the mechanism of conduction in homogeneous solids, in electrolytes and in gases.

We also take the opportunity to look at some sources of e.m.f. related to properties of conducting materials which, together with the electromagnetic induction of chapter 9, completes the survey of sources promised at the end of section 6.1.

12.1 Currents in Extended Media and the Mobility of Carriers

Most of our concern in chapters 6 and 10 has been with currents in metallic wires which are certainly important both practically and fundamentally. Conduction always takes place, however, through a three-dimensional medium and we must develop the appropriate formulae for this. The resistance of a *conductor*, defined in chapter 6 in terms of the potential difference and current between two points in a network, was there contrasted with the resistivity of a *material*:

we now show how resistivity and conductivity can be defined at a point.

Consider a cylindrical element of a current-carrying conductor of conductivity σ as in Fig. 12.1. The potential difference across the ends is $-\mathrm{d}V$ because the potential falls along the direction of positive current flow, and hence

$$-\mathrm{d}V = R\,\mathrm{d}I = \frac{\mathrm{d}l}{\sigma\,\mathrm{d}A}\,\mathrm{d}I$$

by (6.28). Since $-\mathrm{d}V/\mathrm{d}l$ is the electric field strength E in the element and $\mathrm{d}I/\mathrm{d}A$ the current density J, we have in the limit

$$\mathbf{J} = \sigma\mathbf{E} \qquad (12.1)$$

or

$$\mathbf{E} = \rho\mathbf{J} \qquad (12.2)$$

sometimes known as Ohm's law at a point, though they are strictly definitions of σ and ρ at a point. An ohmic conductor would be

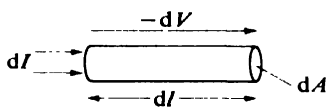


Fig. 12.1. An element of a current-carrying conductor.

one in which σ is independent of the magnitude of $\mathbf{E}$ at constant temperature, etc.

In practice there is a limit to the smallness of the element, set by the interatomic distances in the conductor. As with the definition of density, we can only say that the volume must be small macroscopically but large enough to contain many atoms so that σ is a smooth function of position. (See also the discussion of models in section 13.8.)

A conductor for which σ is the same at every point is *homogeneous* and one for which its value at a point is independent of the direction of $\mathbf{E}$ or $\mathbf{J}$ is *isotropic*. We shall not concern ourselves with non-homogeneous or anisotropic conductors.

Effect of E.m.f. The above only applies when there is no source of e.m.f. in the conductor and the field $\mathbf{E}$ is electrostatic ($\mathbf{E}_Q$ in the notation of section 6.1). If $\mathbf{E}_M$ also exists, $-\mathrm{d}V$ in the derivation of (12.1) must now be replaced by $R\,\mathrm{d}I - \mathrm{d}\mathscr{E}$ from (6.11) and this gives

$$\mathbf{J} = \sigma(\mathbf{E}_M + \mathbf{E}_Q) \qquad (12.3)$$

since the relation between $\mathscr{E}$ and E_M at a point is $E_M = d\mathscr{E}/dl$ from (6.2).

Mobility of Carriers. Let us assume for the moment that the elementary charges carrying the current are all of one type and have a mean drift velocity $\mathbf{v}$ so that by (1.16)

$$\mathbf{J} = ne\mathbf{v} \tag{12.4}$$

where n is the density of carriers (number per unit volume) and e their charge. Thus

$$\sigma = \mathbf{J}/\mathbf{E} = ne\mathbf{v}/\mathbf{E}$$

The *mobility* μ of a carrier is its drift velocity per unit electric field strength so that

$$\mu = \mathbf{v}/\mathbf{E} \qquad \text{(Definition of } \mu) \tag{12.5}$$

and

$$\sigma = ne\mu \tag{12.6}$$

If more than one type of carrier exists then

$$\sigma = \Sigma\, ne\mu \tag{12.7}$$

in which, as in (1.17) and (1.18), the signs of e and μ are included.

12.2 Diffusion of Carriers

If the density of carriers in a conductor is greater at one point than another, diffusion will take place and an electric current will flow in addition to that resulting from the application of an electric field. Such a process is not important in metals because the drift velocities are there so much greater than diffusion velocities, but in semiconductors, ionized gases and electrolytes diffusion plays a significant part in determining the current and is considered briefly here as a result.

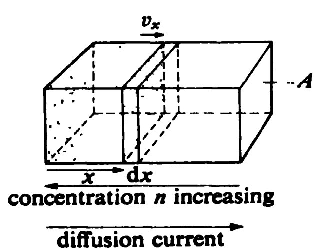


Fig. 12.2. Diffusion of carriers in a conductor.

Suppose the concentration of carriers n decreases as shown in Fig. 12.2 from one end of a part of a conductor to the other. For small concentration gradients Fick's law is assumed to apply, viz. that the number of particles crossing unit area per unit time is proportional to and in the opposite direction to the gradient, or that

$$nv_x = -D\frac{\mathrm{d}n}{\mathrm{d}x} \tag{12.8}$$

where D is the coefficient of diffusion*. This means that there is a current density due to diffusion given by

$$J_x = nev_x = -De\frac{\partial n}{\partial x} \tag{12.9}$$

or in general that $\mathbf{J}$ is $-De$ multiplied by the gradient of the concentration of carriers.

In a conductor where both diffusion and an applied electric field operate in, say, the x-direction

$$J_x = \sigma E - De\,\partial n/\partial x = e(n\mu E - D\,\partial n/\partial x) \tag{12.10}$$

12.3 Metallic Conduction: the Electron Gas Model

The striking features of metallic conduction are the obedience to Ohm's law and to the Wiedemann–Franz law, the latter stating that the ratio of thermal to electrical conductivity is LT where T is the absolute temperature and L a constant (the Lorenz number) which has (almost) the same value for all metals. We have already seen that electrons are responsible for metallic conduction and we therefore adopt an atomic model in which the outermost (or valence) electrons of the metallic atoms are so weakly bound that they move freely amongst the framework formed by the ions, and collide with them (but see the end of section 12.5). Under these circumstances the paths of the electrons are similar to those of the molecules of a gas as conceived in the kinetic theory and a number of kinetic theory concepts are taken over to form an electron gas theory of conduction.

In the absence of an applied electric field let the mean free path of the electrons between collisions be λ and the mean random velocity be $\bar{c}$ so that the mean time between collisions is $\lambda/\bar{c}$ which we shall

* The reader should be able to show, by considering the rate of increase in the number of particles between the planes at x and $x+\mathrm{d}x$ (Anv_x entering, $A(nv_x+\mathrm{d}(nv_x))$ leaving), that $-A\,\mathrm{d}(nv_x)$ = rate of increase of $nA\,\mathrm{d}x$. Using (12.8), this yields $\partial n/\partial t = D\,\partial^2 n/\partial x^2$, a general equation governing diffusion processes subject to Fick's law.

denote by 2τ. When an electric field $\mathbf{E}$ is applied the electrons accelerate in the direction of $\mathbf{E}$ with an acceleration $e\mathbf{E}/m$, thus gaining kinetic energy and a drift velocity in the same direction. We know that overall there is no increase in kinetic energy or drift velocity because a steady current flows, so we must make some assumption about the way this energy is lost. The simplest assumption to make is that at a collision an electron loses all the drift energy gained since the previous collision, the lost energy appearing as vibrations of the framework (heat). We also assume that the drift velocities are much smaller than $\bar{c}$ and that τ is unaffected by the application of $\mathbf{E}$.

Since $e\mathbf{E}/m$ is constant when $\mathbf{E}$ is uniform, the drift velocity just before a collision is $e\mathbf{E}\lambda/m\bar{c}$ (acceleration $\times$ time) and, because the initial velocity is zero and the increase uniform, the mean drift velocity is

$$\mathbf{v} = e\mathbf{E}\lambda/2m\bar{c}$$

and hence from (12.4), assuming that diffusion effects are negligible,

$$\mathbf{J} = \frac{ne^2\lambda}{2m\bar{c}}\mathbf{E}$$

and

$$\sigma = \frac{ne^2\lambda}{2m\bar{c}} = \frac{ne^2\tau}{m} \qquad (12.11)$$

The thermal conductivity K of the same electron gas would be given by

$$K = \tfrac{1}{2}nk\lambda\bar{c} \qquad (12.12)$$

using the usual kinetic theory arguments, k being Boltzmann's constant. From (12.11) and (12.12), $K/\sigma = km\bar{c}^2/e^2$. Now $\bar{c}^2$ is given by $8kT/\pi m$, according to classical kinetic theory, although many authors equate it to $3kT/m$ (which is strictly $\overline{c^2}$). Hence we have

$$\frac{K}{\sigma} = C\left(\frac{k}{e}\right)^2 T \qquad (12.13)$$

where C is $8/\pi$ or 3. Equation (12.11) embodies Ohm's law since none of the quantities on the right would be expected to vary if physical conditions were held constant, and (12.13) predicts the Wiedemann–Franz law with a Lorenz number $C(k/e)^2$ or about $2{\cdot}0 \times 10^{-8}$ V^2/°K^2. Many metals have Lorenz numbers close to

this (e.g. at 100°C, Na 2·19, Al 2·23, Ni 1·83) but others show marked differences (e.g. W 3·20, Bi 2·89). The simple theory runs into greater difficulties when examined in more detail. For instance, the value of σ given by (12.11) when n is assumed to be 1 per atom and λ to be the interatomic distance is of the right order of magnitude at room temperatures but deviates greatly at lower temperatures. Moreover, the electron gas would be expected to contribute an extra $\frac{3}{2}R$ to the molar specific heat of metals in addition to the $3R$ given by the Dulong–Petit law for all solids: in fact, metals differ very little in their specific heats from insulators.

Many of the difficulties are removed by using quantum mechanics (see section 12.5).

12.4 The Hall Effect

If a magnetic field is maintained at right angles to the current flow in a conductor, the equipotentials are no longer planes normal to the original current (Fig. 6.15) but are tilted: this is known as the Hall effect and is revealed by the occurrence of a potential difference V between two points such as A and B in Fig. 12.3a. The Hall coefficient R_H is defined by

$$V = R_H IB/d \quad \text{(Definition of } R_H\text{)} \qquad (12.14)$$

where I is the current, B the magnetic flux density and d the thickness of the conductor parallel to **B**.

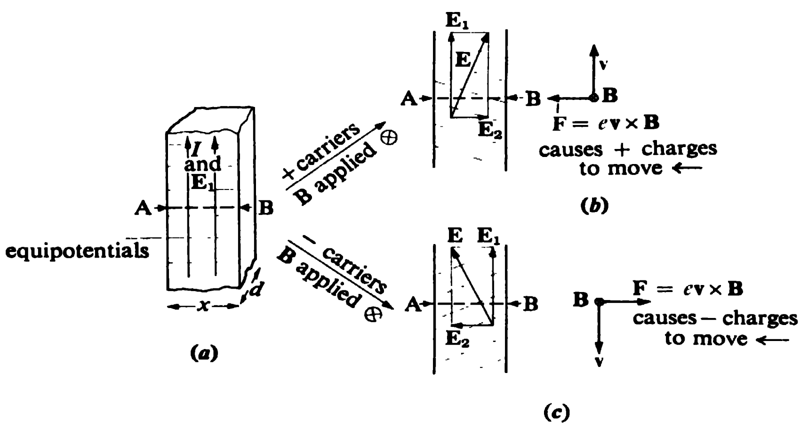


Fig. 12.3. The Hall effect.

The way in which the effect arises is shown in Figs. 12.3b and c for positive and negative carriers respectively. In Fig. 12.3b the velocity of the carriers gives rise to a force $e(\mathbf{v} \times \mathbf{B})$ on them in the direction shown, and this force produces a drift of positive charge to the left leaving the right-hand side negative. These charges accumulate and set up an electrostatic field $\mathbf{E}_2$ which combines with the original $\mathbf{E}_1$ to give a resultant as shown with equipotentials at an angle to their original direction. Accumulation of charge at the sides stops when evB and eE_2 are equal and opposite so that in equilibrium the current flows in its original direction with its original drift velocity $\mathbf{v}$ (note that (12.3) and not (12.1) applies to the lateral motion). If x is the width of the conductor, $E_2 = vB$ and thus $V = vBx$. Since $I = nevxd$ by (1.15),

$$V = BI/ned$$

and hence by (12.14)

$$R_{\mathrm{H}} = 1/ne \qquad (12.15)$$

so that the Hall coefficient gives direct information about the density n of carriers. Figure 12.3c shows that for negative carriers the effect is negative so that the sign of R_{H} gives the sign of the carriers. From (12.6) the mobility $\mu = \sigma R_{\mathrm{H}}$ which can therefore also be obtained.

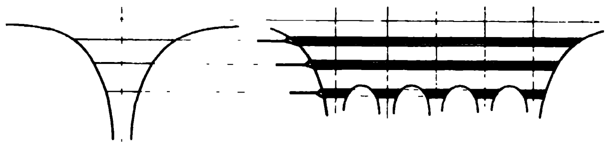
The simple theory is only applicable when one type of carrier is present in a much greater concentration than any other, but it illustrates the principles involved in assessing carrier mobility and concentration.

Without going into the magnitudes of R_{H}, measurements show that while many metals have negative Hall coefficients as expected, some have positive values even when electromechanical experiments have shown that the carriers are electrons (section 6.9—e.g. Mo, Zn, Cd). Some light is thrown on this in the next section.

12.5 The Band Theory of Conduction in Solids

Apart from the positive carriers revealed by the Hall effect, any theory of conduction must explain the properties of semiconductors. These differ from metals not only in their order of magnitude of conductivity, but in having conductivities which increase with rise in temperature and with the addition of small amounts of particular impurities. We can only outline the band theory which explains these effects reasonably well.

Figure 12.4a represents schematically the allowed energies of an electron in the potential field of the nucleus (see appendix 4.2) which can be calculated using quantum mechanics. Quantum theory also shows that when N such atoms are packed together so that they interact strongly, each of the allowed energy levels splits into a

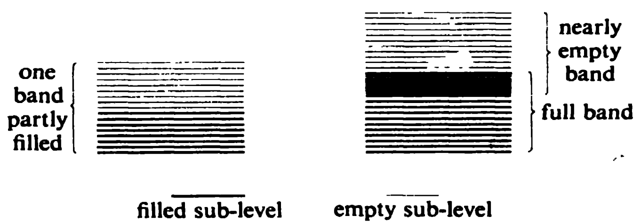


(a) (b)

Fig. 12.4. Production of bands of energy levels in a solid.

band of N sub-levels and, according to Pauli's exclusion principle (which accounts for the periodic structure of the elements) each sub-level may not contain more than two electrons with opposite spins. The essential feature which concerns conduction is that no electron may exist with an energy other than that of a sub-level.

In a metal, the outer electrons forming the electron gas of section 12.3 have sub-levels, at only slightly higher energies, immediately available to them, either because the band is not full (Fig. 12.5a) or because a full band overlaps an empty one (Fig. 12.5b). In these



(a) (b)

Fig. 12.5. Energy bands in a good conductor.

cases conduction takes place when an electric field is applied: the small increase in energy takes the electrons to vacant levels.

In an insulator, the outer electrons fill their band and the next higher one is at such an energy that electric fields of normal magnitudes cannot provide enough to cross the gap (in fact, a few electrons must find their way to the vacant band by thermal excitation and this accounts for the very small but finite conductivity

shown by even good insulators). The uppermost band containing the valence electrons is known as the *valence band* and the next vacant one as the *conduction band*, the gap between them being the *energy gap*.

A semiconductor may be a pure material with a small energy gap (e.g. Si 1·21 eV, Ge 0·78 eV) so that thermal excitation causes enough electrons to leave the valence for the conduction band to give a much higher density of carriers than in an insulator (Fig. 12.6b).

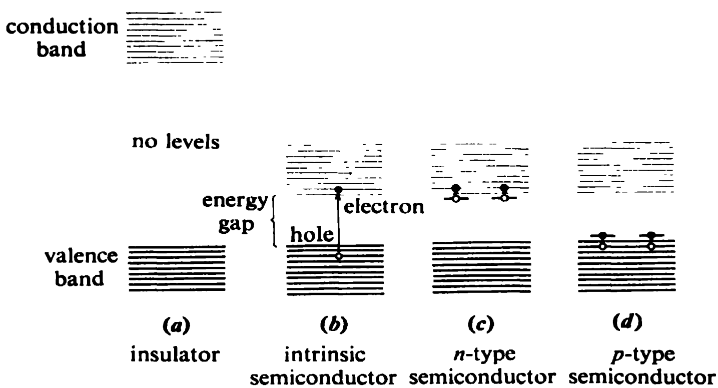


Fig. 12.6. Energy bands in insulators and semiconductors.

This would be known as an *intrinsic* semiconductor. Not only is any current carried by the electrons in the conduction band, but detailed investigation of the motion of the remaining electrons in the valence band shows that the vacancies left may be treated as positive carriers (or *holes*) rather as the motion of the liquid in a spirit-level could be described in terms of the motion of the bubble. Clearly, such conduction will increase as the temperature rises.

In other semiconductors conduction takes place because of the presence of a small concentration of impurity atoms. These may provide electrons in levels (*donor* levels) near the conduction band so that excitation into this band is very probable: conduction then takes place largely by electrons, giving an n-type semiconductor. Alternatively the impurity atoms provide vacant levels (*acceptor* levels) near the valence band so that excitation of an electron from this band is very probable, leaving holes which are largely responsible for the conduction: this forms a p-type semiconductor (Figs. 12.6c and d).

Finally, two points about metals. The positive Hall coefficients in metals such as Zn arise from the fact that their bands are more than half full and thus behave in the metallic framework like the nearly filled bands of a semiconductor—as if they conducted by positive holes. Naturally, the electromechanical experiments which measure the inertia continue to show electrons as the true carriers. The second point concerns the increase in the resistivity of metals with rise in temperature. Once again quantum theory shows that, contrary to the classical picture, a perfectly periodic framework of ions would give zero resistance and that finite values are only obtained because of thermal vibrations (which are responsible for the increase in question) and defects and impurities.

12.6 Conduction in Liquids and Gases

Liquids. The mechanism of conduction in liquid metals is similar to that in solids, but electrolytes, whether fused or in solution, conduct by migration of ions (section 6.9). Salts such as KCl (or $ZnSO_4$) consist of ions K^+ and Cl^- (or Zn^{++} and SO_4^{--}) which attract each other electrostatically at large distances but repel if they are so close that their electron clouds begin to overlap. In the solid state the ions form a regular crystalline array in which the interatomic distances give a minimum potential energy and zero resultant force. The weakening of the electrostatic forces in solvents with high relative permittivity causes many salts to dissociate into their separate ions and thus go into solution. The conductivity of dilute solutions is still given by $\Sigma\, ne\mu$ but, because the ionic masses are much greater than that of an electron, the mobilities and hence the conductivities are much smaller than those of metals.

During electrolysis, the electrode potentials (section 12.8) of the various ions in solution with respect to the material of the electrodes determine which of several possible processes occur.

Gases. The characteristic curve for gaseous conduction shown in Fig. 6.2c is redrawn here as Fig. 12.7. The shape is explained in terms of (*a*) primary ionization due to background radiation (cosmic rays, fallout, etc.) and to any deliberately introduced ionizing source (such as X-rays, α-particles, etc.), (*b*) recombination of positive and negative ions occurring mainly on walls and electrodes and (*c*) secondary ionization, in which primary ions have gained enough kinetic energy from an applied field to cause further ionization by collision.

With no applied field, processes (*a*) and (*b*) are in equilibrium, but

the application of a small potential difference between the values 0 and V_1 causes a small current to flow and reduces (*b*). Between V_1 and V_2 the current is large enough to convey the ions to the electrodes as fast as they are produced by (*a*) and process (*b*) ceases: the saturation current level is thus a function only of the rate of primary ionization. As the electric field increases, more and more ions are accelerated sufficiently for process (*c*) to come into play and increase the current further. At V_3, the breakdown voltage,

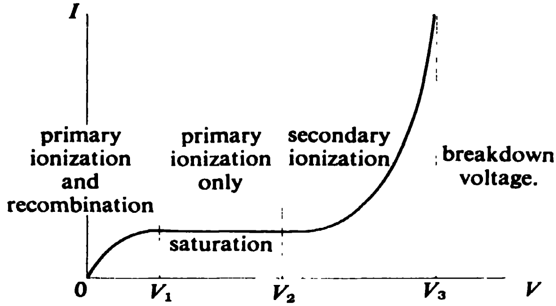


Fig. 12.7. Conduction in gases.

the current increases indefinitely and one of several forms of discharge occurs. *Spark discharge* occurs at normal pressures and is transitory, while *glow discharge*, at low pressures, is maintained by emission of electrons from the cathode by positive ion bombardment or the photoelectric effect. An *arc discharge* on the other hand is maintained by thermionic emission from the cathode which becomes hot by virtue of the high current density used. *Corona discharge* occurs around conductors of small radius of curvature (section 4.3).

Gaseous conduction finds one of its principal applications in devices for counting nuclear particles, which form the primary source of ionization in the above description. Most devices have one electrode in the form of a central wire so that the electric field varies in intensity between the electrodes and the characteristic of Fig. 12.7 is not entirely appropriate. Broadly, however, an *ionization chamber* works in the saturation region between V_1 and V_2, a *proportional counter* between V_2 and V_3 and a *Geiger–Müller tube* above V_3, the discharge produced by a single particle in the last-named being quenched by various means. The characteristic of

Fig. 12.7 also suggests that the gaseous discharge could be used to regulate voltage since the minimum discharge voltage V_3 is a function only of the separation and shape of the electrodes and of the nature and pressure of the gas. Thus the voltage regulator tube in Fig. 12.8 fixes the potential difference across the load at V_3 as the

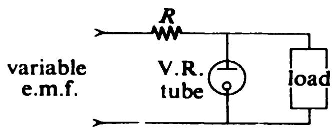


Fig. 12.8. Voltage regulation.

input e.m.f. varies, the current always adjusting so that excess voltage is dropped in R.

12.7 Sources of E.m.f. 1: Contact Potentials and Thermoelectricity

In section 6.1 we saw that e.m.f.s resulted from the existence of non-electrostatic fields $\mathbf{E}_M$ for which $\oint \mathbf{E}_M \cdot d\mathbf{s} \neq 0$. We have already encountered two types: one due to mechanical forces as in the Van de Graaff generator and one due to electromagnetic induction as in A.C. generators, the betatron and the Hall effect. In this section and the next we consider non-electrostatic fields $\mathbf{E}_M$ acting across the boundary between two dissimilar materials: these fields cause charge to flow from one material to the other until the accumulation sets up an opposing $\mathbf{E}_Q$ such that $\mathbf{E}_M + \mathbf{E}_Q = 0$. Figure 12.9 shows how this leads to the formation of an electric double layer at the

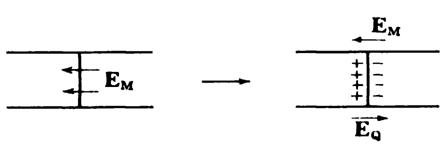


Fig. 12.9. Formation of an electric double layer.

boundary. We consider examples of such effects between two metals in this section and between metal and electrolyte in the next: a further example is the electrification by friction or contact first mentioned in section 1.1.

The origin of E_M in these cases is a matter for electron and atomic theory. One other similar effect can, however, be explained at least qualitatively. A temperature gradient maintained in a single material causes a greater diffusion of electrons from the high to low temperature than vice versa and once again an opposing E_Q is set up and produces equilibrium.

Contact Potentials. If two metals at the same uniform temperature are in contact, electrons are transferred across the junction under the action of an E_M as above: a potential difference is thus set up between the two metals known as the *contact potential.* The transferred charge distributes itself in such a way as not only to oppose E_M but also to produce equipotential volumes over the metals. In the arrangement of Fig. 12.10a X is at a higher potential

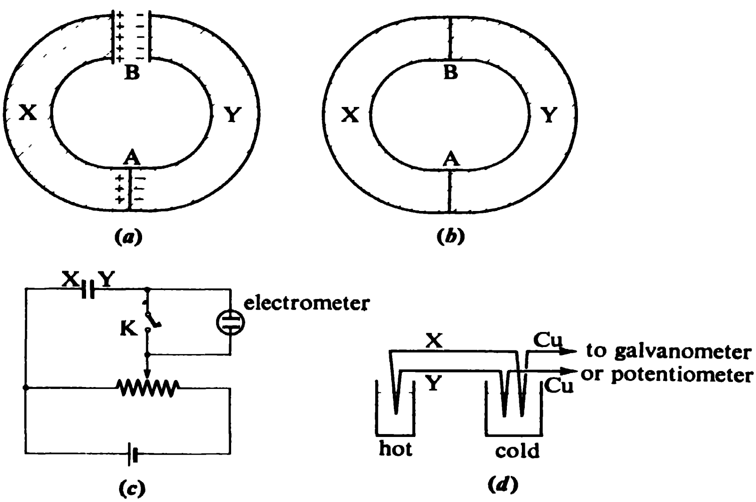


Fig. 12.10. Contact potentials and thermoelectric e.m.f.s.

than Y and a potential difference therefore exists across the gap B equal to the contact potential. When the gap is closed, as in Fig. 12.10b, the two contact potentials are equal if the temperatures of the junctions are the same, and no current can flow in the now completed conducting circuit. If three metals are connected in a ring at the same temperature, the contact potentials V_{12}, V_{23} and V_{31} must be related by $\Sigma V = 0$ (equation (6.13)). It follows that

any junction such as A may be opened and a third metal inserted without upsetting the contact potential at B provided all junctions remain at the same temperature.

Contact potentials may be demonstrated and measured by opening A in Fig. 12.10a and applying a measured potential difference V from a potential divider (Fig. 12.10c with K closed). This changes the potential difference across B by the same amount, and when V is equal and opposite to the contact potential the total potential difference across the gap (drawn as condenser plates) is zero. If the condenser is now isolated by opening K, the absence of potential difference and hence of charge can be detected by seeing whether an electrometer connected across the gap shows any change when the separation of the plates is altered.

Thermoelectric Effects. If a temperature difference is now introduced between the junctions A and B in Fig. 12.10b, the contact potential at A is no longer necessarily equal to that at B. In addition, the temperature gradients in each metal cause further E_M's due to the diffusion of electrons. The result is a total e.m.f. round the circuit known as the *Seebeck e.m.f.*, $\mathscr{E}_{XY}$, which can be measured by opening the circuit of Fig. 12.10b at any point and inserting a potentiometer (Fig. 12.10d): the only condition to be fulfilled is that the two extra junctions introduced and everything between them should be at the same temperature. (It should be emphasized that, if A and B of Fig. 12.10a are at different temperatures, X and Y are no longer equipotential volumes.) The *thermoelectric power* P_{XY} is the rate of change of $\mathscr{E}_{XY}$ with temperature:

$$P_{XY} = d\mathscr{E}_{XY}/dT \quad \text{(Definition of } P_{XY}\text{)} \quad (12.16)$$

and because $\mathscr{E}_{XY}$ in many cases varies quadratically with T, P_{XY} is often approximately a linear function of T.

The pair of metals X and Y form a *thermocouple* whose e.m.f. can be used as a measure of temperature by maintaining one junction (the cold junction in Fig. 12.10d) at a fixed temperature and allowing the other to take up the temperature required. Thermal e.m.f.s in metals are commonly of the order of a few microvolts per °C difference in temperature between junctions, but with semiconductors the effect is much greater and is some few millivolts per °C. (Compare the magnitudes of contact potentials, which are of the order of 1 V.)

Because junctions in electrical networks are heated differentially by the electric currents, undesirable thermal e.m.f.s occur. The usual

method of eliminating the effect is to reverse the main current I. This does not affect the temperatures of the junctions (heating effect $\propto I^2$) and hence leaves the thermal e.m.f.s acting in the same direction in the network. If these thermal e.m.f.s produce currents δI, the main currents are $I+\delta I$ and $-I+\delta I$, by the superposition theorem. In linear deflecting instruments, deflections on opposite sides of the zero are produced whose mean eliminates the effect of δI.

Because the Seebeck e.m.f. drives round the circuit a current which can be made to do work, the energy needed to maintain such a current can only come from the sources of heat maintaining the system at its steady temperatures. We should therefore expect a thermocouple to absorb heat at higher temperatures and reject some of it at lower temperatures. Two effects involving such heat exchanges are known.

In the first place, it is found that a current flowing through a junction between two conductors absorbs or liberates heat according to its direction. This is the *Peltier effect*, the converse of the Seebeck effect. The Peltier coefficient π_{XY} is defined as the rate of absorption of heat per unit current at a junction for current flowing from X to Y.

Secondly, it is found that, in the homogeneous materials connecting the junctions, there is heat produced (in addition to the Joule heat) due to the flow of the electric current up or down a temperature gradient. This effect was predicted by Thomson (later Lord Kelvin) and is known as the *Thomson effect*. The Thomson coefficient σ_T is defined as the rate of absorption of heat per unit current per unit temperature gradient.

The following relations between the three effects can be obtained by thermodynamic reasoning and verified by experiment:

$$P_{XY} = \pi_{XY}/T = \frac{d\pi_{XY}}{dT} + (\sigma_{TX} - \sigma_{TY})$$

$$dP_{XY}/dT = (\sigma_{TX} - \sigma_{TY})/T$$

Many thermodynamic derivations of these relations are of doubtful validity and the reader is recommended to consult the references at the end of the chapter for reliable accounts.

Photovoltaic Effect. Although unconnected with thermoelectricity, it should be mentioned here that, if light irradiates the junction between a p-type and an n-type semiconductor or between a metal and a semiconductor, an e.m.f. is set up which will produce

a small current in an external circuit. This is the photovoltaic effect and the energy needed to sustain the current comes this time from the incident light. Many photocells use this phenomenon.

12.8 Sources of E.m.f. 2: Chemical Cells

When a metal electrode is immersed in an electrolyte, an electric double layer is set up at the interface. This may be due to the transfer of electrons or ions from solution to metal or vice versa: which of the many possibilities occurs in a particular instance depends on the substances used. When equilibrium is reached, a potential difference exists between the metal and the electrolyte known as the *electrode potential*, similar to the contact potential of the last section.

If a second electrode of different material is immersed in the same electrolyte, its electrode potential is usually different from the first, so that a potential difference is set up between the electrodes. We now have a chemical cell whose e.m.f. is equal to the difference between the two electrode potentials and which will supply a current through a circuit formed by connecting the electrodes externally. This time the energy comes from the chemical changes taking place within the cell.

Electrode potentials can only be measured with respect to a standard electrode, invariably hydrogen, so that a complete cell is formed. With this standard, Zn in Zn^{++} has an electrode potential

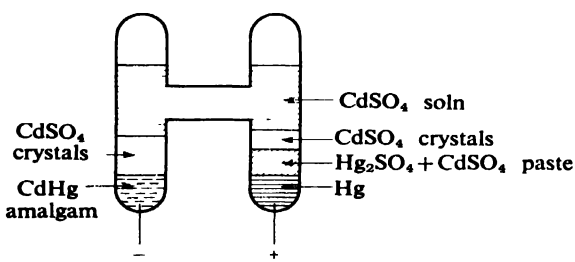


Fig. 12.11. Saturated Weston cadmium cell.

of $-0{\cdot}76$ V and Cu in Cu^{++} has $+0{\cdot}33$ V. A Daniell cell thus has an e.m.f. of approximately 1·09 V. Further details of electrochemical processes can be found in texts on physical chemistry, while details of the construction of primary cells (e.g. the dry battery) and secondary cells (e.g. the lead–acid and Ni–Fe cells) will

be found in more elementary texts. Standard cells are invariably saturated Weston cadmium cells constructed as shown in Fig. 12.11.

12.9 Summary of Chapter 12

Macroscopically, our only advance in this chapter has been to extend the idea of conductivity by defining it at a point through the relation

$$\mathbf{J} = \sigma\mathbf{E} \tag{12.1}$$

We have been concerned mainly with the microscopic mechanisms of conduction in solids, liquids and gases. In general, we showed that

$$\sigma = ne\mu \tag{12.6}$$

where n is the carrier density, e the carrier charge and μ its mobility; and that the Hall coefficient R_H would give a measure of $1/ne$ so that n and μ could be separately determined if one carrier type only were responsible for conduction. A simple electron gas theory gives a mobility of $e\tau/m$ (equation (12.11)) and explains why the conductivity of electrolytes is so low (large mass of the ions) and why the currents in gases are largely carried by electrons rather than positive ions. The band theory of solid conduction removes some of the difficulties of the electron gas theory and accounts for the behaviour of semiconductors.

We also saw that (12.1) had to be extended when non-electrostatic electric fields occurred, to

$$\mathbf{J} = \sigma(\mathbf{E}_M + \mathbf{E}_Q) \tag{12.3}$$

and when diffusion occurred to

$$\mathbf{J} = \sigma\mathbf{E} - De\,\mathbf{grad}\,n \qquad \text{from (12.10)}$$

as long as Fick's law holds, **grad** n being the three-dimensional gradient of concentration whose x-component is $\partial n/\partial x$ as in the original (12.10). (See problem 8.17.)

References

Conduction is treated in more detail in Hemenway, Henry and Caulton (1922) and at a more advanced level in Cusack (1922). Reliable accounts of thermoelectricity are to be found in Cusack and in Zemansky (1922).

PROBLEMS

SECTION 12.1

12.1 Show that the rate of production of heat per unit volume at a point in a conductor is $\mathbf{J}\cdot\mathbf{E}$ or $\sigma\mathbf{E}^2$ where $\mathbf{J}$ is the current density, $\mathbf{E}$ the electric field strength and σ the conductivity.

12.2 Calculate the mobility of electrons in copper assuming that each atom contributes one conduction electron.

12.3 A *neutral* current is defined as one in which the densities of positive and negative carriers, n_p and n_n, are equal so that $n_p e_p + n_n e_n = 0$ because $e_n = -e_p$. Show that the magnitude of a neutral current (e.g. in a copper wire) depends only on the *relative* drift velocity of the two types of carrier and is independent of any motion of the observer.

12.4 A *charged* current is one in which $\rho \equiv n_p e_p + n_n e_n \neq 0$. Show how the magnitude of a charged current depends on the velocity v of an observer O relative to an origin P fixed in the laboratory.

SECTION 12.3

12.5 Use equation (12.11) to determine λ for copper assuming that $\bar{c}$ is given by $(8kT/\pi m)^{1/2}$.

SECTION 12.4

12.6 The Hall angle θ_H is the angle between the direction of current flow and the resultant $\mathbf{E}$ of Fig. 12.3. Show that $\tan\theta_H = \mu B$ where μ is the mobility of the carriers, and show also that θ_H for copper is always very small.

SECTION 12.5

*12.7 Evaluate kT in eV at room temperature, k being Boltzmann's constant and T the absolute temperature. How does the result account for substances with energy gaps of 1 eV having conductivities much lower than those of metals?

CHAPTER 13

DIELECTRIC MATERIALS

No substance is a perfect insulator, but there are many with a conductivity small enough in a steady electric field for it to be neglected as a first approximation. Such substances are known as *dielectric materials*, or more loosely as *dielectrics*, and so far we have two pieces of experimental evidence about their behaviour:

1. The effect of an insulator on capacitance was described in section 5.4 in terms of the relative permittivity ε_r, and in section 13.1 we carry the description of dielectrics as far as we can using this concept only. We shall see, however, that we can deal only with the most restricted examples.

2. The attraction between a charge and a neutral insulator (section 1.1) implied that the positive and negative charges in the latter move small distances in opposite directions. This movement and consequent separation of bound charges by an electric field was described as *polarization* at the end of section 3.8: this concept is developed quantitatively in section 13.2 and later linked with ε_r through electric susceptibility and displacement.

13.1 Relative Permittivity

The relative permittivity ε_r of a substance was defined in (5.14) by

$$C_m/C_0 = \varepsilon_r \quad \text{(Definition of } \varepsilon_r\text{)} \quad (13.1)$$

where C_0 is the capacitance of a condenser *in vacuo* and C_m that of the same condenser with the insulating medium filling the whole of space in which an electric field exists. The ε_r's of actual materials are thus obtained from measurements of capacitance both with and without the medium.

For this section, we confine ourselves to linear, isotropic and homogeneous materials, recognizing linearity by the proportionality of Q and V, isotropy by ε_r being independent of the orientation of the insulator between the plates and homogeneity by ε_r being independent of the particular sample chosen and its arrangement in

the condenser. Such materials are sometimes called class A dielectrics but we shall use the initials LIH to describe them, omitting one or more initials when we wish to be more general. For LIH dielectrics, ε_r is found experimentally to be independent of the shape of the condenser, to be always greater than 1 and to vary with temperature, frequency, etc. (see section 5.4). The value in a steady field is known as the static relative permittivity: most substances show little variation at low frequencies and it is the static value we shall be dealing with for most of the chapter.

Effect of Infinite Medium on Electrostatic Formulae. When the vacuum round a condenser is completely filled with an infinite LIH dielectric the capacitance clearly increases by ε_r. If there are no connections to the plates, Q will remain constant and, because $Q = VC$, the potential difference will fall by ε_r. Since this effect is independent of the shape of the plates, the *potential difference between* and the *potential of* conductors carrying any charge drops by a factor ε_r in the presence of an infinite LIH medium.

It does *not* follow at once that the *potential at any point due to* the charged conductors will also fall, e.g. that because $V = Q/4\pi\varepsilon_r\varepsilon_0 a$ is the potential *of* a charged sphere, $Q/4\pi\varepsilon_r\varepsilon_0 r$ is the potential *due to* the sphere at $r > a$ from the centre. However, we see that the same reduction in potential would be produced if the dielectric were absent and Q were reduced to Q/ε_r. It is therefore clearly possible for the reduction to have occurred because of a surface charge on the dielectric adjacent to the charge on the conductor and of opposite sign (as long as no volume charges occur: see the end of section 13.4). If we assume this generally for LIH dielectrics, it follows that all expressions for V in chapters 3 and 4 must be modified by the addition of ε_r in the denominators:

$$\left.\begin{aligned} V_\sigma &= \iint_\sigma \frac{\sigma\,\mathrm{d}S}{4\pi\varepsilon_r\varepsilon_0 r} \quad \text{(cf. (3.16))};\\ V_{\text{pt. chge}} &= Q/4\pi\varepsilon_r\varepsilon_0 r \quad \text{(cf. (3.13)), etc.} \end{aligned}\right. \qquad (13.2)$$

Electric field strength *defined as negative potential gradient* must also fall everywhere by ε_r and thus in particular:

$$\left.\begin{aligned} \mathbf{E}_{\text{pt. chge}} &= \frac{Q}{4\pi\varepsilon_r\varepsilon_0 r^2}\,\hat{\mathbf{r}} \quad \text{(cf. (3.4))};\\ E_\sigma &= \sigma/\varepsilon_r\varepsilon_0 \quad \text{(cf. (3.6)), etc.} \end{aligned}\right\} \qquad (13.3)$$

We must not, however, assume that **E** is also equal to the force on a small charge in the medium. This is a point discussed later in section 13.8 and it should be noted that, without this assumption, (13.3) cannot be extended to Coulomb's law: it does not necessarily become $F = Q_1Q_2/4\pi\varepsilon_r\varepsilon_0 r^2$ in an infinite LIH medium, much less in any other.

The effective surface charge density on the dielectric will be denoted by σ_P as opposed to the conduction charge density σ_c on the adjacent conductor. By (13.2), σ_P must reduce the field of σ_c to that of σ_c/ε_r and hence σ_P must have a magnitude $(\varepsilon_r - 1)\sigma_c/\varepsilon_r$ and a sign opposite to that of σ_c. Using (13.3),

$$\sigma_P = \varepsilon_0(\varepsilon_r - 1)E_m \tag{13.4}$$

where E_m is the electric field strength in the medium. (See Fig. 13.1.)

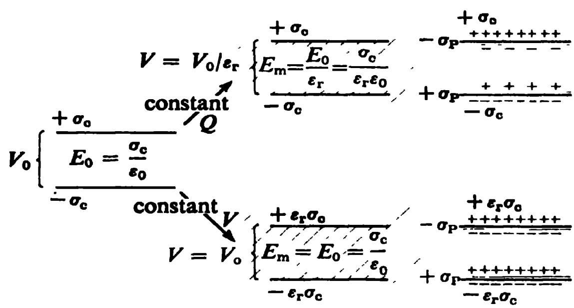


Fig. 13.1. The effect of introducing a dielectric into a condenser.

Introduction of Dielectric at Constant Potential. The definition (13.1) does not depend on the way the dielectric is introduced and still applies if the plates are maintained at constant potential difference by batteries. Now, however, the increase in capacitance must be accompanied by an increase in the conduction charge by a factor ε_r since V in $Q = VC$ is constant. Hence the batteries supply an extra charge density $(\varepsilon_r - 1)\sigma_c$ to the plates. The electric field strength, like V, remains constant and is the same as if the charge density were still just σ_c (see Fig. 13.1). The effective surface charge density on the dielectric, σ_P, is thus $(\varepsilon_r - 1)\sigma_c$ and, because E_m is now σ_c/ε_0, σ_P is again given by $\varepsilon_0(\varepsilon_r - 1)E_m$ as in (13.4).

Figure 13.1 illustrates the two cases for a parallel-plate condenser.

13.2 Polarization and Electric Susceptibility

We now look for an explanation of dielectric behaviour in terms of the polarization which we believe takes place, and we do not now restrict ourselves always to infinite LIH media. In an unpolarized element of an insulator as in Fig. 13.2a, the centroids of the positive and negative charges coincide and no external field is produced. The effect of polarization is shown in Fig. 13.2b, where small movements of the charges have taken place in opposite directions. The polarization **P** at a point is defined in direction as the direction of

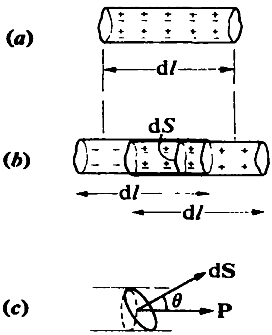


Fig. 13.2. (a) An element of unpolarized material; (b) the displacement of charge in polarized material; (c) the charge crossing dS *is* $P\mathrm{d}S\cos\theta$.

displacement of the positive charges, and in magnitude as the resultant positive charge per unit area crossing an area normal to the displacement. It follows that the charge crossing dS in Fig. 13.2b is $P\,\mathrm{d}S$ and is the sum of the +charge crossing from left to right and the −charge crossing right to left. It also follows (Fig. 13.2c) that the charge crossing an area **dS** whose plane is not normal to **P** will be that crossing $\mathrm{d}S\cos\theta$, i.e. $P\,\mathrm{d}S\cos\theta$ or $\mathbf{P}\cdot\mathrm{d}\mathbf{S}$. We may thus regard as a definition of **P**:

$$\text{Resultant charge crossing an area } \mathrm{d}\mathbf{S} = \mathbf{P}\cdot\mathrm{d}\mathbf{S} \qquad \text{(Definition of } \mathbf{P}\text{)} \quad (13.5)$$

If an element of volume as in Fig. 13.2b is considered in isolation, an unbalanced +charge of magnitude $P\,\mathrm{d}S$ appears at one end and an unbalanced −charge of equal magnitude appears at the other end, as long as the element is small enough for **P** not to vary from

one end to the other. The element thus forms an electric dipole of moment $d\mathbf{p} = \mathbf{P}\, dS\, dl$ or $\mathbf{P}\, d\tau$ where $d\tau$ is the volume. Hence in the limit:

$$\text{Electric dipole moment of volume } d\tau = \mathbf{P}\, d\tau \qquad (13.6)$$

which may be regarded as an alternative definition of **P**. In all cases the element of dielectric, while macroscopically small, must be large enough to contain many atoms.

Surface and Volume Polarization Charges. If **P** is uniform throughout a block of dielectric, the positive charges at one end of an element such as that of Fig. 13.2b will overlap the negative ones at the end of the adjacent element and will be of the same density: the only unbalanced charges will occur at the outer surfaces and these will have a density given by (13.5). Thus

$$\text{Surface density of polarization charges, } \sigma_P = P_n \qquad (13.7)$$

where P_n is the normal component of **P**.

If **P** is not uniform, the charge crossing an area at one end of an element is not necessarily equal in magnitude to that crossing the same area of the adjacent element and the amount of charge in any small region would not remain zero. This would give rise to a volume density of polarization charge ρ_P in addition to the σ_P of (13.7).

The electric field at any point, whether inside or outside the dielectric, is due to all the charges present, and if it is legitimate to consider an insulating medium as consisting of a smoothed-out distribution of charge as we have done so far, then **E** is simply the field produced by any conduction charges present plus that due to ρ_P and σ_P.

Electric Susceptibility. If the electric field strength at a point in a medium is **E** and the polarization **P**, the electric susceptibility χ_e is defined by

$$\mathbf{P} = \varepsilon_0 \chi_e \mathbf{E} \quad (\text{Definition of } \chi_e) \quad (13.8)$$

where the ε_0 is inserted so that χ_e shall be dimensionless. The units of **P** from its definition will be those of charge per unit area, the same as those of $\varepsilon_0\mathbf{E}$. (See note in appendix 13.1.) A dielectric is linear if χ_e is independent of the magnitude of **E**, is homogeneous if it is independent of position and is isotropic if it is independent of the direction of **E**.

13.3 General Electrostatic Laws and Electric Displacement

In section 4.9, we saw that the properties of electrostatic fields were summarized by

$$\oint \mathbf{E}\cdot d\mathbf{s} = 0 \qquad (3.17)=(13.9)$$

and

$$\oiint \mathbf{E}\cdot d\mathbf{S} = \Sigma\, Q/\varepsilon_0 \qquad (4.2)=(13.10)$$

the latter being Gauss's theorem. These are unaffected by the presence of dielectrics if **E** is now given by both conduction and polarization charges so that (13.10) becomes

$$\oiint \mathbf{E}\cdot d\mathbf{S} = \Sigma\, Q_c/\varepsilon_0 + \Sigma\, Q_P/\varepsilon_0 \qquad (13.11)$$

A typical situation is illustrated in Fig. 13.3 in which the Gaussian surface S intersects some dielectrics such as B and C and completely encloses others such as A. Only B and C can contribute

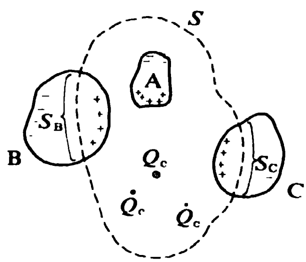


Fig. 13.3. A Gaussian surface with dielectrics present.

to Q_P since the total enclosed charge due to A is zero. The positive polarization charge inside S will be the negative of that which has crossed S_B and S_C from inside S to outside it during the polarization of B and C. By equation (13.5), Q_P inside S will be $-\iint_{(S_B+S_C)} \mathbf{P}\cdot d\mathbf{S}$, the negative sign occurring because of our sign convention attached to areas (appendix 4.1). Since $\mathbf{P}=0$ over the rest of S

$$\Sigma\, Q_P = -\oiint_S \mathbf{P}\cdot d\mathbf{S} \qquad (13.12)$$

and hence from (13.11)

$$\oiint_S (\varepsilon_0\mathbf{E}+\mathbf{P})\cdot d\mathbf{S} = \sum_S Q_c \qquad (13.13)$$

The *electric displacement* $\mathbf{D}$ (unit C/m^2) is defined by

$$\mathbf{D} = \varepsilon_0\mathbf{E}+\mathbf{P} \quad \text{(Definition of } \mathbf{D}) \tag{13.14}$$

and (13.13) becomes

$$\oint\!\!\oint_S \mathbf{D}\cdot d\mathbf{S} = \sum_S Q_c \tag{13.15}$$

This can be regarded as Gauss's theorem in the presence of dielectrics, and it shows that while both conduction and polarization charges are sources of $\mathbf{E}$, only conduction charges are sources of $\mathbf{D}$. The form of (13.15) shows that for distributions of conduction charge considered in chapters 3 and 4, $\mathbf{D}$ is in all cases to be obtained by multiplying by ε_0 all the expressions for $\mathbf{E}$ derived therein, since no dielectrics were present and $\mathbf{P}=0$ everywhere.

'Displacement' is not a happy choice of name: it originated historically from the idea that it was the result of a movement of charge in the dielectric (giving $\mathbf{P}$) and a similar state of strain in space (giving $\varepsilon_0\mathbf{E}$). Note that $\mathbf{D}$ is not a *necessary* quantity in that dielectric behaviour could be described in terms of $\mathbf{E}$ and $\mathbf{P}$, but it is a very convenient one.

Relation between Susceptibility and Relative Permeability. By substituting (13.8) in (13.14):

$$\mathbf{D} = \varepsilon_0(1+\chi_e)\mathbf{E} \tag{13.16}$$

Consider now any system of conduction charges. Whether these are *in vacuo* or in an infinite LIH medium, $\mathbf{D}$ at any point is the same since only Q_c gives rise to it, so that (13.16) tells us that $\mathbf{E}$ in the medium will fall to $1/(1+\chi_e)$ of its vacuum value. This change can be identified with that described in section 13.1 so that

$$\varepsilon_r = 1+\chi_e \tag{13.17}$$

and hence, from (13.16),

$$\mathbf{D} = \varepsilon_r\varepsilon_0\mathbf{E} \tag{13.18}$$

and, from (13.8)

$$\mathbf{P} = \varepsilon_0(\varepsilon_r-1)\mathbf{E} \tag{13.19}$$

(cf. equation (13.4)). The last three equations were strictly only derived for an infinite LIH medium, but we now agree to extend the concept of relative permittivity by using (13.17) as a general definition. This now defines ε_r at a point in a medium where (13.1) could

give no more than a mean value. With this agreement, (13.18) and (13.19) are also general.

Vectors in a Dielectric. Figure 13.4 illustrates the way in which polarization charges reduce the field inside a dielectric, making ε_r always greater than 1. Alternatively, we can regard **D** as arising

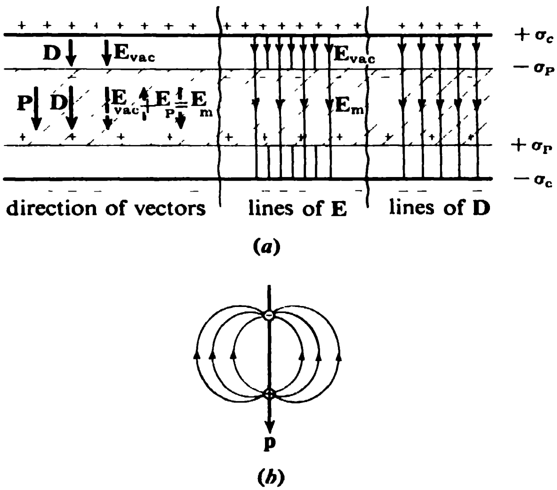


Fig. 13.4. (a) Vectors in a LIH dielectric between the plates of a parallel-plate condenser; (b) elementary dipole showing flux and field in a direction opposite to that of **p**.

only from conduction charges and, since $\varepsilon_0\mathbf{E} = \mathbf{D} - \mathbf{P}$, **E** is reduced in any medium in which **P** and **D** are in the same direction.

13.4 LIH and Non-LIH Dielectrics

None of the relations and definitions in sections 13.2 and 13.3 are limited to LIH dielectrics, but it must be emphasized that χ_e and hence ε_r are not necessarily constants but may vary from one point to another either because of non-homogeneity or because non-linearity occurs and **E** is not uniform: because of this and because ε_r also varies with physical conditions, the term 'dielectric constant' is not favoured. Moreover, if the dielectric is anisotropic, (13.8) is an abbreviation for

$$P_x = \varepsilon_0\chi_{xx}E_x + \varepsilon_0\chi_{xy}E_y + \varepsilon_0\chi_{xz}E_z \tag{13.20}$$

with two similar expressions for P_y and P_z. Equation (13.18) is similar and both χ_e and ε_r are *tensors* rather than scalars: in this

case **D**, **E** and **P** are not necessarily in the same direction. Thus, while (13.18) is sometimes used as a definition of **D** in preference to (13.14), it is deceptively simple: only in LIH dielectrics are χ_e and ε_r scalar constants under constant physical conditions.

Anisotropy is confined to single crystals whose symmetry is tetragonal or lower and to substances which are strained: we shall generally assume isotropy. Non-linearity would occur at very high electric fields when saturation sets in or at low fields in substances known as *ferroelectric* (only by analogy with *ferromagnetic*—ferroelectricity has nothing to do with iron). Non-homogeneity is caused by variations in composition, density and structure of the dielectric.

Polarization Charge in LIH Dielectrics. In section 13.2, a polarized dielectric was seen to be equivalent to a set of surface and volume charges. If we consider any surface S wholly within a polarized LIH material, the polarization charge within S will be given by (13.12):

$$Q_P = -\oint\!\!\oint_S \mathbf{P}\cdot d\mathbf{S}$$

If no conduction charges lie inside S, $\oint\!\!\oint_S \mathbf{D}\cdot d\mathbf{S} = 0$. However, because ε_r and χ_e are independent of position,

$$\oint\!\!\oint_S \mathbf{D}\cdot d\mathbf{S} = \varepsilon_r\varepsilon_0 \oint\!\!\oint_S \mathbf{E}\cdot d\mathbf{S} = \frac{\varepsilon_r}{\chi_e}\oint\!\!\oint_S \mathbf{P}\cdot d\mathbf{S} = -\frac{\varepsilon_r}{\chi_e} Q_P$$

and hence Q_P within S is zero. No volume polarization charges therefore accumulate in an LIH dielectric and the relations of section 13.1 are justified because (13.4) follows from (13.7) and (13.19).

13.5 Boundaries. Condensers with more than one Dielectric

Figure 13.5 shows the surface between two media labelled 1 and 2, either of which may be conductors or insulators. Along this surface there resides in general a conduction charge of surface density σ_c and a polarization charge. The generalized Gauss's theorem (13.15) applied to the volume in Fig. 13.5a generated by the lines of **D** forming the curved sides and having elementary surfaces $d\mathbf{S}_1$ and $d\mathbf{S}_2$ parallel to the surface S gives

$$\mathbf{D}_1\cdot d\mathbf{S}_1 - \mathbf{D}_2\cdot d\mathbf{S}_2 = \sigma_c\, dS$$

As the heights of the cylinders shrink, both $\mathbf{dS}_1$ and $\mathbf{dS}_2$ tend to $\mathbf{dS}$ and in the limit

$$D_{1n}-D_{2n} = \delta D_n = \sigma_c \tag{13.21}$$

where D_n is the component of $\mathbf{D}$ normal to S. Thus we can say that in general *the normal component of* $\mathbf{D}$ *is discontinuous by* σ_c *across any surface.*

In Fig. 13.5b, equation (13.9) is applied to a closed path consisting of two elements $\mathbf{ds}_1$ and $\mathbf{ds}_2$ as shown and completed by two small

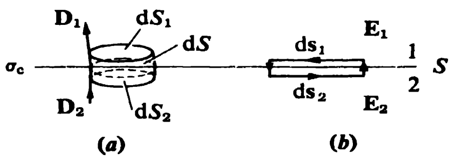


Fig. 13.5. Boundary conditions.

portions at the ends. The contributions of the ends to $\oint \mathbf{E}\cdot\mathbf{ds}$ vanishes in the limit as long as $\mathbf{E}$ is finite, so that

$$E_{1t}\,\mathrm{d}s_1 - E_{2t}\,\mathrm{d}s_2 = 0$$

and hence, because $\mathrm{d}s_1 = \mathrm{d}s_2$,

$$E_{1t} = E_{2t} \qquad \text{or} \qquad \delta E_t = 0 \tag{13.22}$$

where E_t is the component of $\mathbf{E}$ tangential to the surface. Thus *the tangential component of* $\mathbf{E}$ *is continuous across any surface.*

Condensers with more than one Dielectric. An application of (13.21) and (13.22) occurs in condensers where the space between the plates contains more than one dielectric, the bounding surfaces being either perpendicular or parallel to the lines of force.

When the surfaces are perpendicular to the lines of $\mathbf{D}$ and $\mathbf{E}$, (13.21) shows that $\mathbf{D}$ will be the same on both sides of any one surface while $\mathbf{E}$ must therefore change inversely as the ε_r's. In fact the values of $\mathbf{E}$ and therefore of the potential differences between the surfaces of the dielectrics must each be $1/\varepsilon_r$ of the values they would have *in vacuo*. If the potential differences are V_1, V_2, etc. the total capacitance C will be $Q/(V_1+V_2+\cdots)$ and hence $1/C = 1/C_1 + 1/C_2 + \cdots$ where $C_1 = Q/V_1$, etc., showing that C may be obtained by treating the single condenser as a set in series.

Figure 13.6 illustrates the case of a parallel-plate condenser with

a dielectric slab of thickness t. For this the capacitance is given by

$$\frac{1}{C} = \frac{x-t}{\varepsilon_0 A} + \frac{t}{\varepsilon_r \varepsilon_0 A}$$

or

$$C = \frac{\varepsilon_0 A}{x - t(1 - 1/\varepsilon_r)}$$

showing that the effect of the slab is the same as that of reducing the distance apart of the plates *in vacuo* by $t(1 - 1/\varepsilon_r)$.

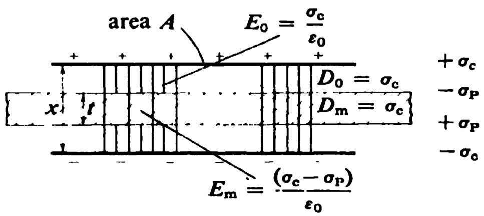


Fig. 13.6. Boundary of dielectric perpendicular to lines of **D** *and* **E**.

A similar argument for boundaries parallel to the lines of force shows that capacitance may be obtained by treating the condenser as a set in parallel: Fig. 13.7 shows this for a parallel-plate condenser. Note that in both Figs. 13.6 and 13.7, the conduction charge density

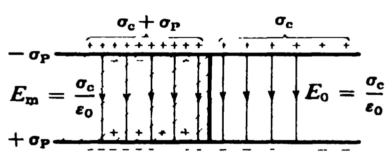


Fig. 13.7. Boundary of dielectric parallel to lines of **D** *and* **E**.

σ_c is not necessarily that which the plates would possess if the dielectric were withdrawn, since what then happens depends on whether or not the plates are isolated: the figures show merely a static situation.

13.6 Electric Energy in the Presence of Dielectrics

In sections 5.5 and 5.6 it was shown that the work done in assembling a set of charges could be expressed as $\Sigma \frac{1}{2}QV$ or

$\iiint \frac{1}{2}\varepsilon_0 E^2 \, d\tau$ in the absence of dielectrics. Similar arguments can be used to obtain more general expressions but some simplifying assumptions must be made.

Firstly, while we may not always wish to insist on linearity, ε_r must be single-valued so that any integrals have a unique magnitude: our treatment will not necessarily apply when hysteresis occurs in ferroelectrics.

Secondly, we must treat systems including dielectrics as thermodynamic rather than purely mechanical so that instead of equating work done on a system W to the increase in mechanical energy U, we should specify the conditions more precisely. For instance, for an adiabatic change, $dW = dU$, the increase in internal energy, while for a reversible isothermal change, $dW = dF$, the increase in Helmholtz free energy. Because ε_r varies with temperature and we wish to avoid this complication, we assume that all processes are carried out reversibly and isothermally so that strictly $dW = dF$: we shall retain the symbol U_E instead of F since it is still electric in origin and plays the same part as the corresponding quantity in chapter 5.

Finally, stresses in a dielectric caused by polarization will normally strain the material (electrostriction) and cause changes in ε_r: again we wish to avoid this and shall assume that all fluids are incompressible.

Work Done in Charging a Condenser and a Conductor. As in section 5.5, the work done by an external agency in transferring a charge dq from one plate of a condenser to the other is, in the limit,

$$dW = v \, dq \tag{13.23}$$

where v is the potential difference. The total work in charging to Q is

$$W = \int_0^Q v \, dq \tag{13.24}$$

and if all dielectrics are linear, $q = vC$ at all stages and

$$U_E = W = \tfrac{1}{2}QV = \tfrac{1}{2}CV^2 = \tfrac{1}{2}Q^2/C \tag{13.25}$$

as before, Q being the conduction charge. The same applies to an isolated conductor.

Energy of Charged Conductors in terms of Q and V (Linear dielectrics). For a general collection of charges we cannot use quite the same argument as in section 5.5 since Coulomb's law does not give the complete force on a conduction charge when dielectrics

are present. Consider, however, three conduction charges Q_1, Q_2 and Q_3 and let the potential at the point Q_1 due to Q_2 alone be V_{12} with corresponding potentials for all the others. Any dielectrics are assumed to remain permanently in position. The work done in assembling Q_1 and Q_2 only is given either by Q_1V_{12} or by Q_2V_{21}, which must therefore be equal to each other and to $\frac{1}{2}Q_1V_{12}+\frac{1}{2}Q_2V_{21}$. In a similar way the addition of Q_3 will entail work $Q_3V_{31}+Q_3V_{32}$ which can also be written as four terms. The total work is thus

$$\begin{aligned} W &= \tfrac{1}{2}Q_1(V_{12}+V_{13})+\tfrac{1}{2}Q_2(V_{21}+V_{23})+\tfrac{1}{2}Q_3(V_{31}+V_{32}) \\ &= \tfrac{1}{2}Q_1V_1+\tfrac{1}{2}Q_2V_2+\tfrac{1}{2}Q_3V_3 \end{aligned}$$

where V_1 is the potential at Q_1 due to all the other charges, etc. This can be generalized for any number of charges to

$$U_E = W = \sum_i \tfrac{1}{2}Q_iV_i \tag{13.26}$$

and, as in section 5.5, to a collection of conductors:

$$U_E = W = \sum_A \tfrac{1}{2}Q_AV_A \tag{13.27}$$

Energy in Terms of Field Quantities. The argument of section 5.6 can be taken over in the presence of LIH dielectrics in which the directions of **D** and **E** are the same at any point. Referring back to Fig. 5.18 we now apply the generalized Gauss's theorem to the volume bounded by dQ, dS and the lines of force and obtain $D\,dS=dQ$ so that

$$dU_E = \tfrac{1}{2}D\,dS\,V$$

The relation $dQ=D\,dS$ applies to all the equipotentials cut by the tube and thus

$$dU_E = \tfrac{1}{2}D\,dS\int_{S'}^{\infty} E\,dr = \int_{S'}^{\infty} \tfrac{1}{2}DE\,dS\,dr$$

Hence the work or free energy is

$$U_E = \iiint \tfrac{1}{2}DE\,d\tau = \iiint \tfrac{1}{2}\varepsilon_r\varepsilon_0E^2\,d\tau \tag{13.28}$$

Further, equation (13.23) gave us the work done in increasing the amount of charge by dq at a potential v, and by following this through the same derivation to (13.28) we see that this is also

$$dW = \iiint \tfrac{1}{2}E\,dD\,d\tau \tag{13.29}$$

13.7 Forces and Changes in Energy

As long as there is no contact between conductors and dielectrics the expressions for forces on *conductors* obtained in section 5.7 are still valid since the same arguments can be taken over. The forces on any *dielectrics* can also be obtained from the energy. For instance, at constant Q, if a dielectric has a force $\mathbf{F}$ acting on it internally and is kept in equilibrium by an external force $\mathbf{G} = -\mathbf{F}$, the same arguments can be used to obtain (5.22) and thus (5.23) for the dielectric in a condenser:

$$F_s = -\left(\frac{\partial U_E}{\partial s}\right)_{Q,\,D} = \frac{1}{2}V^2\frac{\partial C}{\partial s} \tag{13.30}$$

and any external work done on the system results in an equal increment of the free energy.

At constant V, as before,

$$F_s = +\left(\frac{\partial U_E}{\partial s}\right)_{V,\,E} = \frac{1}{2}V^2\frac{\partial C}{\partial s} \tag{13.31}$$

and of the energy supplied by the battery, half goes in doing external work and half in the increase of free energy (see problem 13.9).

Equation (13.30) shows that the force on a dielectric is always such as to try to decrease U_E when the conductors are isolated and when both the conduction charges Q and the displacement $\mathbf{D}$ are constant. If U_E is written in the form $\iiint \frac{1}{2}D^2/\varepsilon_r\varepsilon_0 \, d\tau$, a given dielectric will clearly tend to move into regions of larger $\mathbf{D}$ because this reduces the value of the integrand, ε_r being always greater than 1. This is a generalization of the attraction between a point charge and a dielectric body first met in section 1.1.

Force on the Surface of a Charged Conductor. If there is no contact between dielectric and conductor the analysis is the same as in section 5.7 and the outward force is $\sigma_c^2/2\varepsilon_0$ per unit area. When a solid dielectric touches a charged conducting surface the force is indeterminate because the form of contact is indefinite. Only for a fluid can a calculation be performed, and this case we now discuss.

If we use the energy method of section 5.7, we again find that the force per unit area is equal to U_E outside the conducting surface, i.e.

$$\text{outward pressure on the surface of a charged conductor in contact with a fluid dielectric} = \tfrac{1}{2}DE = \sigma_c^2/2\varepsilon_r\varepsilon_0 \tag{13.32}$$

since $E = \sigma_c/\varepsilon_r\varepsilon_0$ and $D = \sigma_c$.

This result can be confirmed for a parallel-plate condenser by using (13.30). There is, on the other hand, the first method of section 5.7: if there were a small gap left between the fluid dielectric and the conducting surface, that method would still yield an outward pressure of ${\sigma_c}^2/2\varepsilon_0$. How is it that the mere contact of the fluid changes this? An analysis of the forces acting on a polarized dielectric reveals that there are two sources of hydrostatic pressure acting outwards from the dielectric surface which reduce the pressure on the conductor to (13.32). The first is due to a force on every element of volume $d\tau$ given by $P\,d\tau\,dE/ds$, using (4.26), where **s** is the direction of both **P** and **E**: it can be written as $\frac{1}{2}\varepsilon_0\chi_e(\partial E^2/\partial s)$ per unit volume and it produces a hydrostatic pressure gradient of the same value (Fig. 13.8a). There is thus an excess pressure inside a polarized dielectric of $\frac{1}{2}\varepsilon_0\chi_e E^2$ compared with a region where $E=0$ (Fig. 13.8b). The second source of pressure is the unbalanced

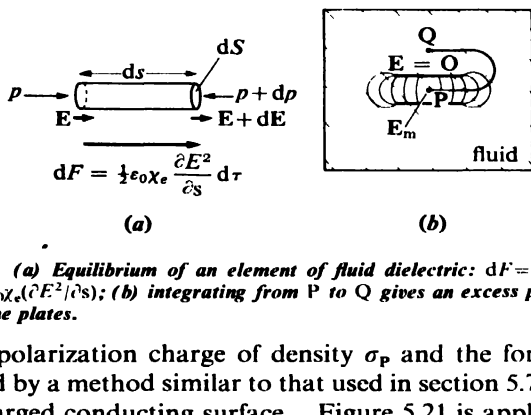


Fig. 13.8. (a) Equilibrium of an element of fluid dielectric: $dF=dp\,dS$ *and hence* $dp/ds=\frac{1}{2}\varepsilon_0\chi_e(\partial E^2/\partial s)$*; (b) integrating from* P *to* Q *gives an excess pressure* $\frac{1}{2}\varepsilon_0\chi_e E^2$ *between the plates.*

surface polarization charge of density σ_P and the force on this is obtained by a method similar to that used in section 5.7 for the force on a charged conducting surface. Figure 5.21 is applicable except that the inside field is E_m and the outside field E_0. The integration of $\varepsilon_0 E'\,dE'$ is, however, between the limits 0 and E_0-E_m because a force due to E_m acts on the negative ends of the dipoles whose positive ends are in the surface layer. The outward pressure is thus $\frac{1}{2}\varepsilon_0(E_0-E_m)^2$ and, because $E_0=\varepsilon_r E_m$ and $\sigma_P=P=\varepsilon_0(\varepsilon_r-1)E_m$, this becomes $P^2/2\varepsilon_0$. Since both forces oppose the original $\sigma^2/2\varepsilon_0$, the resultant pressure is

$$\frac{D^2}{2\varepsilon_0}-\frac{\varepsilon_0\chi_e E^2}{2}-\frac{P^2}{2\varepsilon_0}$$

using $D=\sigma_c$. This reduces to (13.32), surprisingly.

13.8 Electric Fields within Charge Distributions

In chapter 4 we shelved the question of what was meant by the electric field strength **E** within a distribution of charge, and the same problem arises in dielectric theory: we ask how, if at all, **E** is related to the force per unit charge inside a dielectric.

It is important to realize that calculations of fields due to, and forces on, charge distributions are performed on a *model* and the results compared with experiment. Good agreement means that the model is a satisfactory one as far as those experiments go, while complete disagreement means that the model must be discarded (cf. the Thomson and Rutherford atoms in section 4.3). It more often happens, however, that different models are required to account for different sets of experimental results or that a crude model is sufficient for one set but that a more detailed model must be used for another. Thus we found in chapter 5 that our model of a charged conductor as a body carrying a continuous distribution of charge on its surface was quite adequate to account for all the results connected with capacitance, although we realized that other experiments require us to assume the existence of discrete charges.

When we are only interested in macroscopic or large-scale phenomena it is often possible to adopt a perfectly adequate model in which only continuous distributions of charge exist: the reason for this was discussed in section 1.5. Only one thing about this need worry us: even if we have only continuous distributions and no point charges, is it not possible that **E** and V become infinite within such distributions? The expressions (4.12) and (4.13) for points within a spherical distribution show that, in this case at any rate, **E** and V converge to finite values. In general, (4.9) shows that V at the centre of a spherical shell of charge, radius r, thickness $\mathrm{d}r$ and volume density ρ, is $\rho r\,\mathrm{d}r/\varepsilon_0$ which tends to zero as r tends to zero, while **E** is zero at the centre for any r. Thus the immediately adjacent charges of such a continuous distribution do not cause **E** and V to become infinite.

The next question concerns the meaning of **E**. We know that *outside* conductors and dielectrics, the models we have adopted of smoothed-out distributions are adequate because we can measure **E** and confirm our results. *Inside* conductors and insulators, however, we have an apparent contradiction: on the one hand we have that **E** in conductors is zero and in dielectrics is the field due to external sources plus the gradient of the V given by the equivalent

σ_P and ρ_P: these fields we shall call the macroscopic electric field. On the other hand, we know that matter is discrete and that **E** *defined as force per unit charge* within any material will fluctuate violently from point to point and from time to time: this we shall call the microscopic field, often denoted by **e**.

We should expect the macroscopic field to be the space and time average of **e**, such as might be experienced by a charge moving rapidly through the material. Another way of imparting physical significance to the macroscopic field in a dielectric is to realize that, if we scoop out a cavity within the material which is macroscopically small but large enough for a point within it to be counted as *outside* the dielectric, we shall obtain a force per unit charge which does not fluctuate like **e**. However, the force now depends on the shape of the cavity since extra polarization charges will appear on its inner surface.

If such a cavity is needle-like with its axis in the direction of **P** then the polarization charges will appear at the ends and can be made negligibly small (Fig. 13.9a). The force per unit charge is

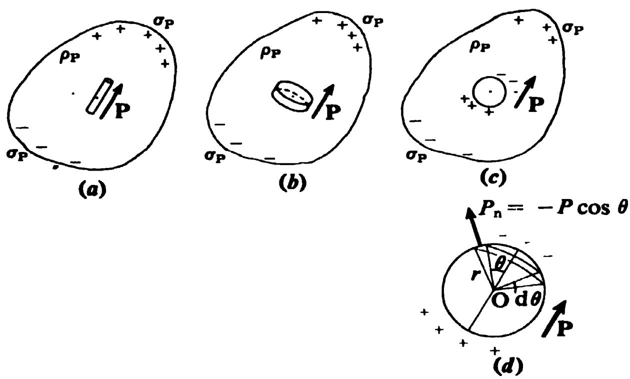


Fig. 13.9. Cavities in dielectrics.

then due only to σ_P, ρ_P and any external field, and is thus just **E**. A disc-like cavity with its plane perpendicular to **P** as in Fig. 13.9b carries a surface density of polarization charge equal to $\pm\mathbf{P}$ on its flat faces which will produce an extra force $\mathbf{P}/\varepsilon_0$ per unit charge by (4.16): the total force per unit charge in such a cavity is thus $\mathbf{E}+\mathbf{P}/\varepsilon_0$ or $\mathbf{D}/\varepsilon_0$.

In a spherical cavity of radius r, the polarization charges appear-

ing on the inner surface are of surface density $-P_n$ or $-P\cos\theta$ (Figs. 13.9c and d). According to problem 3.3, a ring of charge like the elementary strip in the figure produces an electric field at O of $(Q\cos\theta)/4\pi\varepsilon_0 r^2$ where Q is the total charge—in this case $(-P\cos\theta \times 2\pi r\sin\theta)r\,d\theta$. Thus the field due to the elementary strip is $-P\cos^2\theta\sin\theta\,d\theta/2\varepsilon_0$ and due to the whole surface is

$$\int_{\theta=0}^{\theta=\pi} \frac{-P\cos^2\theta}{2\varepsilon_0}\,d(\cos\theta) = \frac{P}{3\varepsilon_0}$$

and the total force in the cavity is

$$\mathbf{E}+\frac{\mathbf{P}}{3\varepsilon_0} \text{ per unit charge} \tag{13.33}$$

If we enquire further as to the field experienced by an actual molecule within the dielectric we are asking for a microscopic theory and this we consider below in section 13.9.

The Status of Coulomb's Law in a Dielectric. Mention must be made of the 'law' $F=Q_1Q_2/4\pi\varepsilon_r\varepsilon_0 r^2$ purporting to give the force between point charges in a medium and sometimes used to define ε_r. This is a complex matter which can only be properly discussed from a more advanced viewpoint, but if we split it into its two components $E=Q/4\pi\varepsilon_r\varepsilon_0 r^2$ and $\mathbf{F}=Q\mathbf{E}$ we can see that it will be of restricted validity. The first component only applies in an infinite LIH medium (sections 13.1 and 13.4), while the discussions in this section above and in the following section are relevant to the validity of $\mathbf{F}=Q\mathbf{E}$: for it depends on what we mean by Q. If we mean it to be the limiting case of a small charged conductor immersed in a fluid, we must take account of the extra forces giving rise to hydrostatic pressure; while if we mean it to be an elementary charge or atom, we must find $\mathbf{F}$ by methods to be discussed in the next section. It will be clear that $F=Q_1Q_2/4\pi\varepsilon_r\varepsilon_0 r^2$ can give little more than an order of magnitude in most cases and should never be used as a basic law, while the view that the force between charges is in some way modified by the medium is misleading.

13.9 The Approach to Microscopic Theory

We know from evidence outside the field of electricity that, on an atomic scale, insulators are built up from groups of atoms or ions forming what we shall call a molecule, although strictly this term should be kept for a more restricted group of materials. Thus, in an inert gas our molecule is a single atom, in a gas such as carbon

dioxide it is a single CO_2 group, while in a solid such as potassium chloride we shall take it as a single KCl group even though the structure is an ionic framework in which molecules do not exist.

Two specifically microscopic quantities are introduced: firstly, the local field **F** experienced by a single molecule, and secondly, the polarizability α defined as the mean electric dipole moment per molecule per unit field. The electric dipole moment of a molecule in a polarized dielectric is thus $\alpha\mathbf{F}$ and because there are $N_A\rho/M$ molecules per unit volume (problem 1.9), it follows from (13.6) that

$$\mathbf{P} = \frac{N_A \rho \alpha}{M}\mathbf{F} \tag{13.34}$$

which relates the microscopic quantities to a macroscopic **P**. Because measurements of ε_r are usually made, it is better to use (13.19) and obtain

$$\varepsilon_r = 1 + \frac{N_A \rho \alpha}{\varepsilon_0 M}\frac{\mathbf{F}}{\mathbf{E}} \tag{13.35}$$

Dielectric theory is concerned with the value of **F**/**E**, because only then can α be calculated and information about the molecules be obtained.

The Local Field. Only for very dilute gases should we expect **F** to be equal to **E**. One method of taking into account the nearby molecules is due to Lorentz: imagine a macroscopically small sphere described about the molecule in question so that all the material outside it gives the field (13.33) and all the molecules inside it must be treated individually and their effect added to (13.33). Lorentz showed that for material with a cubic lattice, the contribution of the nearby molecules would be zero and that it would be very small for fluids in which the random motion produced isotropy. For these materials (13.35) becomes

$$\varepsilon_r = 1 + \frac{N_A \rho \alpha}{\varepsilon_0 M}\left(1 + \frac{\varepsilon_r - 1}{3}\right)$$

using $\mathbf{F} = \mathbf{E} + \mathbf{P}/3\varepsilon_0$ and $\mathbf{P} = \varepsilon_0(\varepsilon_r - 1)\mathbf{E}$. This gives

$$\alpha = \frac{3\varepsilon_0 M}{N_A \rho}\frac{(\varepsilon_r - 1)}{(\varepsilon_r + 2)} \tag{13.36}$$

known as the Clausius–Mossotti formula. This approximation is found to be accurate enough in many cases for α to be calculated from ε_r.

Origin of α. There are three ways in which α may arise, two involving distortion and one orientation. If the molecule has no permanent dipole moment of its own it is said to be *non-polar*, but in an electric field in which the positive and negative charges move in opposite directions, distortion of the molecule will induce a moment. This could be due to the relative motion of nuclei and electrons (*electronic* polarization) or of positive and negative ions in a solid (*ionic* polarization): see Figs. 13.10a and b. If in addition

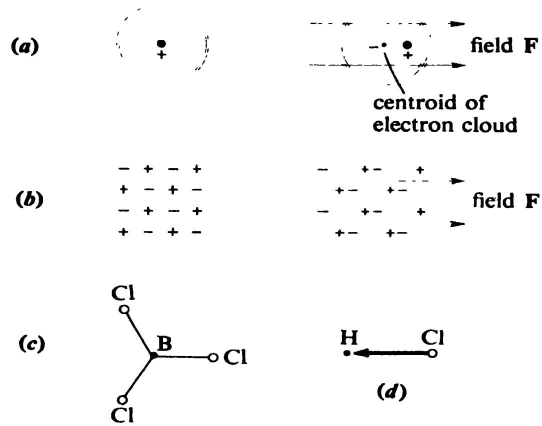


Fig. 13.10. (a) Electronic polarization; (b) ionic polarization; (c) BCl_3*, a non-polar molecule; (d)* HCl*, a polar molecule.*

the molecules are polar, an applied field will tend to turn them into its own direction, an effect which produces an overall polarization (*orientational*) which is temperature-dependent since thermal agitation normally randomizes the directions of the dipoles. Thus if α varies with temperature, the structure is known to be polar (Figs. 13.10c and d) and, though qualitative, this is valuable information.

Problems 13.10 and 13.11 show that the electronic polarizability of a single nuclear atom should be of the order of $3\varepsilon_0 \times$ its volume, and experimental measurements confirm this.

13.10 Dielectric Measurements and Properties

Measurements of relative permittivity below microwave frequencies use the A.C. bridge and resonance methods described in

section 10.12 to obtain C_m/C_0 (equation (13.1)). Microwave methods are outside the scope of this book. At higher frequencies, in the infra-red and visible regions of the electromagnetic spectrum, the relation between ε_r and the refractive index (section 15.5) can be used.

Variation of ε_r with Frequency. Complex Permittivity. Table 13.1 gives some values of dielectric properties for common materials

Table 13.1

DIELECTRIC PROPERTIES OF SOME REPRESENTATIVE MATERIALS

Gases	*Relative permittivity at S.T.P.*	*Dielectric strength*
Hydrogen	1·00027	~2·0 kV/mm
Oxygen	1·00053	~2·6 kV/mm
Dry air	1·00058	~3·0 kV/mm
Carbon dioxide	1·00099	~2·9 kV/mm

*Liquids and Solids** (20°C)

	Relative permittivity at low frequencies	*Power factor*	*Conductivity* (mho/m)	*Dielectric strength* (kV/mm)
Polythene	2·3	0·0002	$10^{\ 11}$	20
Paper	3·7	0·009	$10^{\ 10}$	16
Mica	6·0	0·0002	10^{-14}	100
Ethyl alcohol	26	Large	$3 \times 10^{\ 4}$	—
Water	81	Large	$2 \times 10^{\ 4}$	—
Barium titanate†	4000	0·02	—	

* Approximate values only—materials vary in properties from specimen to specimen in many cases.

† Anisotropic. The quoted ε_r is that for a single crystal along the *a*-axis.

The relative permittivities of gases are taken from Kaye and Laby (1959).

and Fig. 13.11 illustrates the type of variation of ε_r with frequency encountered in a polar material. The high value at low frequencies is the result of orientation polarization. The rotation of the polar molecules into the direction of an applied static field is such that the polarization increases exponentially and a relaxation time or time constant for the process can thus be defined (just as for a *CR* circuit). In an oscillating field, the molecules will thus never be quite in phase with the reversals in direction and at high frequencies will not follow them at all: the orientational contribution falls off.

In a condenser with an ideal dielectric, we have seen that $\mathbf{I}$ and $\mathbf{V}$ are in quadrature and no loss of energy occurs. From (10.30):

$$\mathbf{I} = j\omega\varepsilon_r C_0 \mathbf{V} \tag{13.37}$$

where C_0 is the vacuum capacitance. For a variety of reasons, a real dielectric passes a component of current in phase with $\mathbf{V}$ (e.g. due to conduction current or the dielectric relaxation time just mentioned), and a real term must be added to (13.37). This can be allowed for, whatever the cause, by using a complex relative permittivity $\varepsilon_r{}^*$ equal to $\varepsilon_r' - j\varepsilon_r''$. Thus (13.37) becomes

$$\mathbf{I} = j\omega\varepsilon_r{}^* C_0 \mathbf{V} = j\omega\varepsilon_r' C_0 \mathbf{V} + \omega\varepsilon_r'' C_0 \mathbf{V} \tag{13.38}$$

The ε_r'' component is responsible for the total loss of energy in the dielectric and $\tan\delta = \varepsilon_r''/\varepsilon_r'$ is a measure of this.

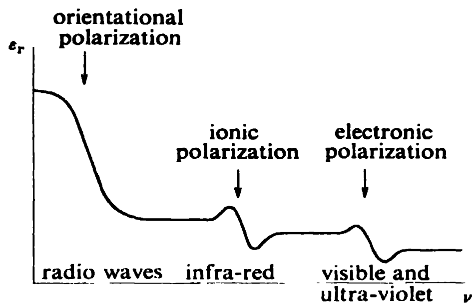


Fig. 13.11. Variation of relative permittivity with frequency.

Other Dielectric Properties. A small class of materials exhibit *ferroelectricity* in which a graph of $\mathbf{D}$ against $\mathbf{E}$ has the same general shape as the ferromagnetic hysteresis loops in the next chapter. These and other substances are capable of permanent polarization in the absence of an electric field, thus forming *electrets* (cf. the word 'magnet'), whose properties are complex. (See Gross, 1944.)

Some anisotropic materials polarize on the application of a mechanical stress and also show mechanical strain when polarized by an electric field: this phenomenon is known as *piezoelectricity* and is distinct from electrostriction (section 13.6). Quartz is the most important example of a piezoelectric material and a properly cut crystal will polarize along one direction if compressed along another at right angles. Mechanical vibrations can be converted into electric ones in this way by incorporating the quartz in a condenser and the sharp mechanical resonance which occurs can be used to control the frequency of electronic oscillators. Some piezoelectric materials also exhibit *pyroelectricity*: polarization

accompanying a change in temperature. Fuller descriptions of all these properties are given in the references cited below.

13.11 Summary of Chapter 13

By introducing the electric displacement **D** we have been able to summarize the electrostatic laws by

$$\oint \mathbf{E}\cdot d\mathbf{s} = 0 \quad \text{and} \quad \oiint \mathbf{D}\cdot d\mathbf{S} = \Sigma\, Q_c \qquad \text{(13.9) and (13.10)}$$

where $\mathbf{D} = \varepsilon_0\mathbf{E}+\mathbf{P}$ and **P** is defined as the displaced charge per unit area or the dipole moment per unit volume ((13.5) and (13.6)). These laws could be summarized by saying that there are no vortices of **E** and that only conduction charges are sources of **D**.

Dielectrics themselves are described by the vectors **P** and **E** between which there exists the general relation

$$\mathbf{P} = \varepsilon_0\chi_e\mathbf{E} \qquad (13.8)$$

defining electric susceptibility which is a scalar constant only if the material is LIH. If we define

$$\varepsilon_r = 1+\chi_e \qquad (13.17)$$

in general, the quantity has the same properties as those of relative permittivity defined in (13.1) by C_m/C_0.

The extension to time-varying fields must wait until chapter 15, but in A.C. circuits the effects of non-ideal dielectrics can be allowed for by assuming a complex ε_r.

Appendix 13.1 Definitions of D and χ_e

The substitution $\varepsilon_0 \to 1/4\pi$ used in chapters 1 to 5 to convert formulae to the CGS e.s.u. form does not work for relations involving **D** and χ_e. The reason for this is that the definitions differ by a factor of 4π:

$$\mathbf{D}_{\mathrm{MKSA}} = \varepsilon_0\mathbf{E}+\mathbf{P} \quad \text{whereas} \quad \mathbf{D}_{\mathrm{CGS}} = 4\pi(\varepsilon_0\mathbf{E}+\mathbf{P})$$

$$\chi_{e\mathrm{MKSA}} = \mathbf{P}/\varepsilon_0\mathbf{E} \quad \text{whereas} \quad \chi_{e\mathrm{CGS}} = \mathbf{P}/4\pi\varepsilon_0\mathbf{E}$$

Hence the substitutions needed to convert all formulae in this chapter to CGS e.s.u. are

$$\varepsilon_0 \to 1/4\pi; \qquad \mathbf{D} \to \mathbf{D}/4\pi; \qquad \chi_e \to 4\pi\chi_e \qquad (13.39)$$

so that, for instance

$$\mathbf{D} = \mathbf{E}+4\pi\mathbf{P}$$

$$\varepsilon_r = 1+4\pi\chi_e$$

$$\mathbf{P} = \chi_e\mathbf{E}, \text{ etc.}$$

References

For further general reading on microscopic dielectric theories, see Cusack (1921) or von Hippel (1922). Birks's annual publication —Progress in Dielectrics—is recommended for recent developments in materials and theory.

PROBLEMS

SECTION 13.1

13.1 A 100 pF parallel-plate air condenser is charged to a potential difference of 50 V. The space between the plates is then completely filled with an insulating liquid of $\varepsilon_r = 3$, a 50 V battery being permanently connected across the plates. Find the new charge. If the filling had been carried out with the plates isolated from the battery, what would the new charge and potential difference have been?

13.2 The condenser of problem 13.1 is connected in parallel with one of 200 pF, the combination is charged to a potential difference of 50 V and the battery is removed. If the first condenser is now filled with the insulating liquid, what are the final charges and potential differences?

SECTION 13.2

13.3 A LIH dielectric sphere has a uniform polarization P. Find an expression for the surface density of polarization charge σ_P at any point on the surface. Show that the 'depolarizing' field produced at the centre of the sphere by σ_P is $P/3\varepsilon_0$.

SECTION 13.5

13.4 A rectangular box has two opposite faces of conducting material so that it forms a parallel-plate condenser. If it is one-third filled with an insulating liquid of $\varepsilon_r = 4$, find the ratio of the capacitance when the conducting faces are horizontal to that when they are vertical. (Neglect edge effects.)

13.5 A long cylindrical condenser with radii a and $4a$ has its inner conductor covered with a cylindrical sleeve of dielectric whose relative permittivity is 2 and whose outer radius is $2a$. Find the capacitance per unit length.

13.6 Show that lines of **E** and of **D** in crossing a plane boundary from one LIH dielectric of relative permittivity ε_1 to a second with ε_2 are refracted; and that the law governing the refraction is $\varepsilon_1/\varepsilon_2 = \tan\theta_1/\tan\theta_2$, where θ_1 and θ_2 are the angles between the lines and the normal to the boundary.

SECTION 13.6

13.7 The plates of a parallel-plate condenser are of area A, distance apart x and are at a difference of potential V. A slab of dielectric of uniform thickness t and relative permittivity ε_r is inserted between the plates. Find the change in electric energy of the condenser if the plates are (*a*) isolated, (*b*) maintained at their initial potential difference by a battery.

13.8 A long cylindrical condenser has radii a and b ($b > a$) and the space between the cylinders is filled with a dielectric of relative permittivity ε_r and dielectric strength K. What is the maximum energy which can be stored per unit length of the condenser?

SECTION 13.7

13.9 A slab of dielectric is of such shape and size as to fill the space between the rectangular plates of a parallel-plate condenser except for negligible air gaps. The slab is withdrawn in a direction parallel to one side of the plates whose length is a. Find the force on the dielectric when only part of it is still between the plates, if the potential difference is maintained at V throughout and if the vacuum capacitance of the condenser is C.

SECTION 13.9

13.10 Orders of magnitude of atomic polarizability can be obtained from simple models as in this problem and the next. Assume that an atom consists of a nucleus with charge $+Q$ and an electron with charge $-Q$ orbiting in a circle of radius a centre at the nucleus. Show that if an electric field **E** is applied in a direction perpendicular to the plane of the orbit and displaces the nucleus by an amount small compared with a, the induced dipole moment is $4\pi\varepsilon_0 a^3\mathbf{E}$, and the polarizability $4\pi\varepsilon_0 a^3$.

13.11 If the atomic model is taken instead as a nucleus of charge $+Q$ at the centre of a spherical volume of radius a carrying the total electronic charge $-Q$, show that the atomic polarizability is again $4\pi\varepsilon_0 a^3$.

SECTION 13.10

*13.12 An electret is in the form of a thin sheet with **P** perpendicular to the faces. Show that if edge effects are neglected, **D** is zero everywhere.

CHAPTER 14

MAGNETIC MATERIALS

The effects of material media on the magnetic laws developed in chapters 7, 8 and 9 bear some resemblance to the effects of dielectrics described in the last chapter and many similarities in treatment will occur. There are, however, complicating factors, the two main ones being (*a*) the existence of the small but technically important class of ferromagnetic materials with their permanent magnetism and hysteresis which require separate treatment (sections 14.6 and 14.7) and (*b*) the existence of two possible models for the elementary dipoles constituting a magnetized material: one based on the monopole and the other on the current loop.

14.1 Relative Permeability

The self-inductance of a circuit *in vacuo* was shown in chapter 9 to depend only on the geometry, but we find experimentally that the presence of material media affects it. The *relative permeability* μ_r of a medium is defined for the moment by

$$L_m/L_0 = \mu_r \quad \text{(Definition of } \mu_r\text{)} \quad (14.1)$$

where L_0 is the self-inductance of a circuit *in vacuo* and L_m that of the same circuit with the medium filling the whole of space in which a magnetic field exists: effectively an infinite medium.

Because the magnetic flux $\Phi = LI$, it follows that if a current I gives a flux linkage Φ_0 *in vacuo* and Φ_m in an infinite medium

$$\Phi_m/\Phi_0 = \mu_r \quad (14.2)$$

as well, and the mean flux density $\bar{B}$ across any circuit also changes by μ_r:

$$\bar{B}_m/\bar{B}_0 = \mu_r \quad (14.3)$$

and both (14.2) and (14.3) are alternative definitions of μ_r.

Definition (14.3) is the basis of one method of measuring μ_r by using a toroidal coil wound on an annular ring of the material, the

flux being confined almost entirely to the inside of the coil (see section 14.10) but only for ferromagnetics is μ_r sufficiently different from 1 to make the method practicable. Other experiments using non-uniform fields, also described in section 14.10, show that all substances are magnetic and enable μ_r's to be determined. These measurements show that there are three main classes of material:

(*a*) *diamagnetic*, with μ_r slightly less than 1 and linear,
(*b*) *paramagnetic*, with μ_r slightly greater than 1 and linear,
(*c*) *ferromagnetic*,* with μ_r much greater than one, non-linear and dependent on previous history. We also distinguish for the moment ideally *soft* ferromagnetics, which retain no magnetism when there is no external magnetic field, from *hard* ferromagnetics, which may form permanent magnets.

Single crystals may be anisotropic, with different μ_r's in different directions, and non-homogeneity may also occur, but *we shall restrict the treatment entirely to isotropic homogeneous* (*IH*) *materials.*

Effect of Infinite Medium on Magnetic Formulae. The definitions of μ_r tell us that when the vacuum round a circuit in which a current I is maintained constant is completely filled with a medium, the inductance, flux and mean flux density across the circuit are all increased by μ_r. But it cannot be deduced that the flux density *at any point* due to these currents will also change by the same factor. However, we see that the same changes would occur if the magnetic material were absent and the currents increased to $\mu_r I$. It is therefore possible for the increased flux density to be the result of currents flowing in the surfaces adjacent to I (as long as no volume currents also occur). If we assume this generally for all IH media of infinite extent, all the expressions for $\mathbf{B}$ in chapters 7 and 8 which contain I must be modified by the addition of a μ_r in the numerator:

$$\left.\begin{aligned} B_{r\,\text{dipole}} &= \mu_r\mu_0 2IA\cos\theta/4\pi r^3 \qquad (7.6)\\ B_{\infty\,\text{solenoid}} &= \mu_r\mu_0 nI \qquad (8.7)\\ d\mathbf{B}_{\text{element}} &= \mu_r\mu_0 I\, d\mathbf{l}\times\mathbf{r}/4\pi r^3 \qquad (8.16),\ \text{etc.}\end{aligned}\right\} \qquad (14.4)$$

* Two further classes may be distinguished: (*d*) *antiferromagnetics*, with μ_r of the same order as paramagnetics but with other properties similar to ferromagnetics; (*e*) *ferrimagnetics* which are similar to class (*c*) but are non-metallic and possess high resistivities: ferrites are mixed oxides of iron belonging to this class. Consideration of (*d*) and (*e*) is beyond the scope of this book (see references at the end of the chapter).

The effect on the monopole formulae of chapter 7 is discussed in section 14.8. While the formulae of (14.4) give **B** measured as a flux density, they do not necessarily give the force per unit current element or the couple per unit dipole within the medium: this will be dealt with in section 14.9.

The parallel between this section and section 13.1 should be noted. We are now suggesting that the effect of an infinite IH magnetic medium is to be found by replacing it with surface currents just as we represented the effect of an infinite dielectric by surface charges. We shall call the original currents producing the fields *conduction currents* and denote them by I_c and their surface density by J_{sc} (section 1.7); while those replacing the material we shall call *magnetization* or *Amperian* currents denoted by I_M and J_{sM}. The ideal solenoid of Fig. 14.1 illustrates the concepts. The conduction current I_c flows in n turns per unit length, so that

$$J_{sc} = nI_c \tag{14.5}$$

and the flux density *in vacuo* can be written as $B_0 = \mu_0 J_{sc}$. With the medium, the flux density increases to $B_m = \mu_0\,\mu_r J_{sc}$ as if the total

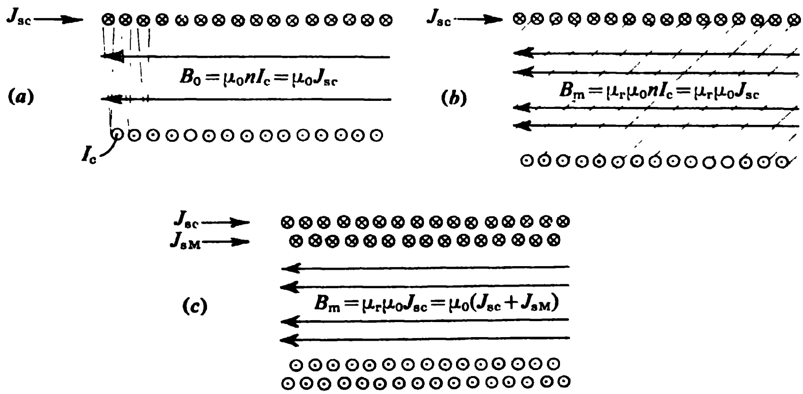


***Fig. 14.1.** (a) Ideal solenoid in vacuo; (b) ideal solenoid with magnetic core; (c) surface currents on the core producing the same flux density as in (b).*

surface current were $\mu_r J_{sc}$. This would be produced as in Fig. 14.1c if the original conduction current J_{sc} were supplemented by a surface Amperian current of

$$J_{sM} = (\mu_r - 1)J_{sc} = \frac{(\mu_r - 1)}{\mu_r}\frac{B_m}{\mu_0} \tag{14.6}$$

14.2 Magnetization

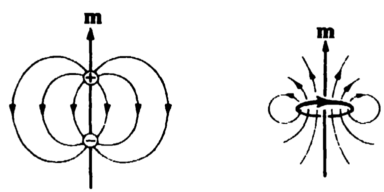
We now look for a model of a magnetic material which will enable us to explain its macroscopic behaviour. Here we are guided by elementary experiments with pieces of soft iron which behave as magnets when placed in a magnetic field and thus can be considered (like a polarized dielectric) as a collection of elementary dipoles. Each small volume $d\tau$ of a magnetized material will possess a magnetic dipole moment $d\mathbf{m} = \mathbf{M}\, d\tau$ where

$\mathbf{M}$ = magnetic dipole moment per unit volume

(Definition of $\mathbf{M}$) (14.7)

defines the *magnetization*, whose unit will be the A/m. The volume $d\tau$ is subject to the same restrictions as that in (13.6).

The polarization $\mathbf{P}$ was interpreted as arising from small displacements of the bound charges whose existence we inferred in dielectrics. We have two alternatives for the origin of $\mathbf{M}$: one that it is due to small displacements of monopoles (cf. $\mathbf{P}$), the other that it arises from the alignment of current loops. No distinction between the two arises if points outside the medium are considered since we have seen in chapter 7 that the flux density at some distance from them is identical. Figure 14.2 shows, however, that the flux of the permanent dipole is in the opposite direction to its moment while



(*a*) (*b*)

Fig. 14.2. Contrast between the internal fields of a two-pole and a current-loop dipole.

that of the current loop is in the same direction, so that the explanation of the observed flux changes on magnetization offered by the two models is bound to differ. We prefer the current-loop model because there is evidence that magnetism is associated with angular momentum which must arise with any current and because it accounts naturally for the observed changes in $\mathbf{B}$. We shall see that the monopole model is the more natural when changes in the magnetic field strength $\mathbf{H}$ are to be explained.

Surface and Volume Currents. A small element of material in isolation, with a length $\mathbf{dl}$ in the direction of $\mathbf{M}$ and a cross-section dS, is shown in Fig. 14.3. If the element is small enough for $\mathbf{M}$ not to vary over it, the equivalent set of current loops have a uniform surface current density J_s and hence a total current $J_s\,dl$. The magnetic moment, by (7.1), is thus $J_s\,\mathbf{dl}\,dS$ but is also, from the definition of $\mathbf{M}$, equal to $\mathbf{M}\,dl\,dS$. Thus for an element in the direction of $\mathbf{M}$

$$\text{Surface density of Amperian currents} = |M| \qquad (14.8)$$

If $\mathbf{M}$ is uniform throughout a large block of material, the Amperian currents on the surface of one element will be equal and opposite to those on the adjacent ones and thus will produce no external

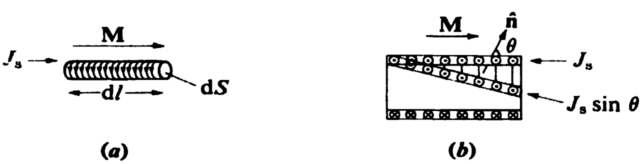


Fig. 14.3. Amperian surface currents.

magnetic effect. Only the currents on the outer surface of the block will remain unbalanced and these will have the value given by (14.8) if the surface has the same direction as $\mathbf{M}$ or if the normal is perpendicular to $\mathbf{M}$. If the normal makes an angle θ with $\mathbf{M}$ then Fig. 14.3b shows that the surface density of Amperian current is $M \sin\theta$. This is what was denoted in section 14.1 by J_{sM} so that

$$\text{Surface density of Amperian currents, } J_{sM} = M\sin\theta = |\mathbf{M}\times\hat{\mathbf{n}}| \qquad (14.9)$$

$\hat{\mathbf{n}}$ being a unit vector normal to the surface and θ the angle between $\hat{\mathbf{n}}$ and $\mathbf{M}$.

If $\mathbf{M}$ is not uniform, the J_{sM} from one element will not balance that from the adjacent one and a volume distribution of Amperian currents J_M may also be necessary.

The magnetic flux density $\mathbf{B}$ at any point, whether inside or outside the material, is due to all the currents present, and if our model is legitimate, $\mathbf{B}$ is simply the flux density due to any conduction currents I_c present plus that due to I_M arising from both J_{sM} and J_M.

14.3 General Magnetic Laws and Magnetic Field Strength

In section 8.8 we saw that the properties of magnetic flux density could be summarized by the circuital theorem

$$\oint_C \mathbf{B}\cdot d\mathbf{s} = \mu_0 I \qquad (8.10)=(14.10)$$

and Gauss's theorem

$$\oiint_S \mathbf{B}\cdot d\mathbf{S} = 0 \qquad (7.26)=(14.11)$$

These are unaffected by the presence of magnetic materials if **B** now results from both conduction currents and Amperian currents, so that (14.11) is exactly the same and (14.10) becomes

$$\oint_C \mathbf{B}\cdot d\mathbf{s} = \mu_0 I_c + \mu_0 I_M \qquad (14.12)$$

Apply this to the typical path C shown in Fig. 14.4 which links conduction currents and passes through pieces of magnetized material like A. Contributions to the right-hand side of (14.12) are

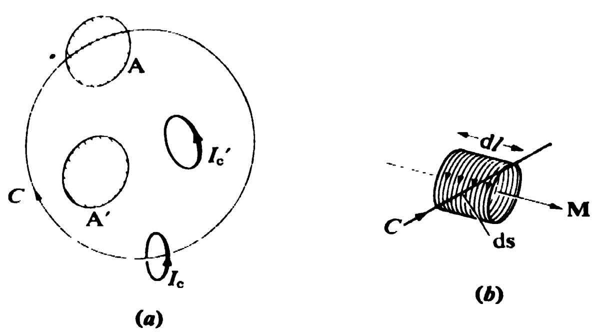


Fig. 14.4. Derivation of the generalized circuital theorem.

only made by currents linked by C so that I_c' and A′ will not contribute. The contribution of A to I_M can be obtained by considering Fig. 14.4b in which **ds** is an element of the path C at a region of A where the magnetization **M** makes an angle α with **ds**. The Amperian currents linked by this part of C are those in the element

of material shown, amounting to $J_s\,dl$. Thus, using (14.8),

$$dI_M = J_s\,dl = M\,ds\cos\alpha = \mathbf{M}\cdot d\mathbf{s}$$

and I_M of (14.12) is $\oint_C \mathbf{M}\cdot d\mathbf{s}$, so that

$$\oint_C (\mathbf{B}/\mu_0 - \mathbf{M})\cdot d\mathbf{s} = I_c \tag{14.13}$$

The *magnetic field strength* $\mathbf{H}$ (unit A/m) is defined by

$$\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M} \quad \text{(Definition of } \mathbf{H}) \tag{14.14}$$

or $\mathbf{B} = \mu_0(\mathbf{H}+\mathbf{M})$. Equation (14.13) becomes

$$\oint_C \mathbf{H}\cdot d\mathbf{s} = I_c \tag{14.15}$$

known as the generalized circuital theorem. In section 7.6 the m.m.f. was defined *in vacuo* as the line integral of $\mathbf{B}\cdot d\mathbf{s}/\mu_0$: we now generalize this to

$$\text{m.m.f. } \mathscr{H} = \oint \mathbf{H}\cdot d\mathbf{s} \quad \text{(Definition of } \mathscr{H}) \tag{14.16}$$

which includes (7.25) as a special case and allows us to quote the circuital theorem as 'the m.m.f. round a closed path is equal to the total conduction current linked'.

In terms of the language introduced in section 4.9, (14.11) shows that there are no *sources* of $\mathbf{B}$ while (14.15) shows that only conduction currents are *vortices* of $\mathbf{H}$ and (14.12) that both conduction and Amperian currents are vortices of $\mathbf{B}$. Great care must be taken over (14.15), however, because it does not tell us anything about *sources* of $\mathbf{H}$: indeed, by substituting $\mathbf{B} = \mu_0(\mathbf{H}+\mathbf{M})$ into (14.11)

$$\oiint_S \mathbf{H}\cdot d\mathbf{S} = -\oiint_S \mathbf{M}\cdot d\mathbf{S} \tag{14.17}$$

which means that the sources of $\mathbf{H}$ are the sinks of $\mathbf{M}$.

Magnetic Susceptibility. The magnetic susceptibility at a point, χ_m, is defined by

$$\mathbf{M} = \chi_m \mathbf{H} \quad \text{(Definition of } \chi_m) \tag{14.18}$$

and is thus dimensionless (but see appendix 14.1). For an IH medium χ_m is independent of position and of the direction of $\mathbf{H}$ and is thus a scalar. Only for linear media is it independent of the magnitude of $\mathbf{H}$.

For an *infinite LIH medium* substitution of $\mathbf{M}=\chi_m\mathbf{H}$ into (14.17) shows that there are no sources of **H** or of **M** since χ_m can be taken outside the integrals. It then follows that **H** can only arise from conduction currents according to (14.15) and is unaffected by the presence of the medium, whereas we saw in section 14.1 that **B** increases μ_r times. The form of the generalized circuital theorem shows that for the distributions of conduction current considered in chapter 8 the magnetic field strength is given by the same expressions as for **B** but divided by μ_0 and that these are unaffected by infinite LIH media. Thus:

$$\left.\begin{aligned} H_{r\ \text{dipole}} &= 2IA\cos\theta/4\pi r^3 \quad \text{from (7.5)} \\ H_{\infty\ \text{solenoid}} &= H_{\text{toroid}} = nI \quad \text{from (8.7)} \\ d\mathbf{H}_{\text{element}} &= I\,d\mathbf{l}\times\mathbf{r}/4\pi r^3 \quad \text{from (8.16) etc.} \end{aligned}\right\} \tag{14.19}$$

Relation between χ_m *and* μ_r. Substituting $\mathbf{M}=\chi_m\mathbf{H}$ in $\mathbf{B}=\mu_0(\mathbf{H}+\mathbf{M})$:

$$\mathbf{B} = \mu_0(1+\chi_m)\mathbf{H} \tag{14.20}$$

This shows us that when, as above, **H** is unaffected by the presence of an LIH medium, **B** increases by a factor $(1+\chi_m)$ so that

$$\mu_r = 1+\chi_m \tag{14.21}$$

and

$$\mathbf{B} = \mu_r\mu_0\mathbf{H} \tag{14.22}$$

Equation (14.21), like (13.17), is an extension of the original definition of μ_r which is now defined at any point in a medium. We see that diamagnetic materials will have small negative susceptibilities and paramagnetic materials small positive ones (table 14.2, p. 379).

Justification of Relations in Section 14.1. Because there are no sources of **M** in an infinite LIH medium there will be no unbalanced volume Amperian currents and only surface currents need be used to replace the medium. Using (14.14) and (14.22)

$$\mathbf{J}_{sM} = \mathbf{M} = \mathbf{B}/\mu_0-\mathbf{H} = \mathbf{B}(1-1/\mu_r)/\mu_0$$

which is the same as (14.6).

14.4 Boundaries and Finite Media

It cannot be too strongly emphasized that, while $\mathbf{B}=\mu_r\mu_0\mathbf{H}$ is a universal relation, it does not imply that **H** is independent of the presence of media and that **B** increases μ_r times—this is something we have so far shown only for *infinite* LIH media.

At any boundary we shall have in general both a conduction surface current J_{sc} and an Amperian surface current J_{sM}. Applying Gauss's theorem $\oint \mathbf{B}\cdot d\mathbf{S}=0$ to the cylinder of Fig. 14.5a as we did for equation (13.21), we obtain the rule that

$$B_{1n}-B_{2n} = \delta B_n = 0 \tag{14.23}$$

where B_n is the component of $\mathbf{B}$ normal to S. Thus *the normal component of* $\mathbf{B}$ *is continuous across any surface.*

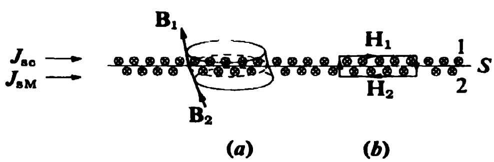


Fig. 14.5. Boundary conditions.

If the generalized circuital theorem $\oint \mathbf{H}\cdot d\mathbf{s}=I_c$ is applied to the path in Fig. 14.5b as we did for equation (13.22), we obtain

$$H_{1t}-H_{2t} = \delta H_t = J_{sc} \tag{14.24}$$

where H_t is the component of $\mathbf{H}$ tangential to S. Thus *the tangential component of* $\mathbf{H}$ *is discontinuous by* J_{sc} *across any surface.*

Boundaries Parallel to Lines of $\mathbf{H}$. When a piece of magnetic material is placed in the field of conduction currents so that it does not fill the whole of the region in which there is a flux density we must first find whether there are any sources of $\mathbf{H}$. In cases where the lines of $\mathbf{H}$ are parallel to any boundaries, $\mathbf{H}$ is continuous across them by (14.24) and hence there are no sources. $\mathbf{H}$ is therefore determined only by its vortices according to the circuital theorem and $\mathbf{B}$ by $\mu_r\mu_0\mathbf{H}$.

As an example consider the ideal solenoid in Fig. 14.6 through the centre of which a rod of cross-section a ($<A$) and of infinite

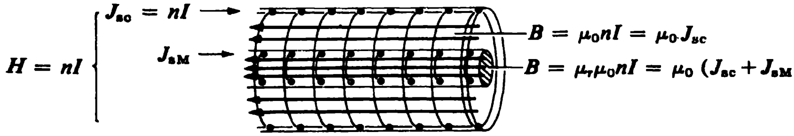


Fig. 14.6. Self-inductance of an ideal solenoid with a magnetic core.

length is inserted. The magnetic field strength is nI everywhere inside the solenoid so that in the material $B=\mu_r\mu_0nI$ while in the space between the material and the solenoid $B=\mu_0nI$. The flux linkage per turn is thus $\mu_r\mu_0nIa+\mu_0nI(A-a)$. The self-inductance is then easily seen to be $\mu_0n^2[A+a(\mu_r-1)]$ per unit length.

Boundaries Perpendicular to Lines of **B**. In this case **B** is continuous across the boundaries by (14.23) and **H** must therefore jump by a factor $1/\mu_r$. There are therefore sources of **H** at such boundaries, as is otherwise clear because lines of **M** must begin and end at them to form sources and sinks (Fig. 14.7a). We shall see that for permanent magnets it is useful to interpret these sources as poles,

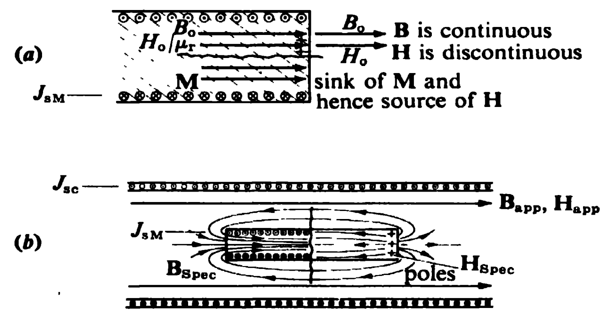


Fig. 14.7. (a) A boundary perpendicular to B *and* H; *(b) sources and vortices of* B *and* H *in a magnetized rod. For clarity, the fields of the solenoid and of the specimen are drawn separately: the resultants are obtained by superposition.* M *in the specimen is in the same direction as* B.

but, for para-, dia- and soft ferromagnetic materials, even apparently simple examples are difficult to solve. While uniform magnetization is not strictly possible, consider a long thin cylinder inside an ideal current-carrying solenoid (Fig. 14.7b). It will be approximately true that **M** and hence **H** and **B** will be approximately uniform within such a rod. The total flux density is the sum of that due to the solenoid, $\mathbf{B}_{app}$, and that due to the Amperian currents in the surface of the rod, $\mathbf{B}_{spec}$. The magnetic field strength will be the sum of $\mathbf{H}_{app}$ and **H** due to sources located at the ends of the specimen, $\mathbf{H}_{spec}$. It does not matter, for points outside the specimen, whether the **B** field or **H** field is calculated, but inside the material we see that if χ_m is positive $\mathbf{H}_{spec}$ opposes $\mathbf{H}_{app}$. For para- and diamagnetics, $\mathbf{H}_{spec}$ is so small that the resultant **H** is in the

same direction as the resultant **B**: this is not so in ferromagnetics, as we see later.

14.5 Magnetic Energy and Forces

The same conditions must apply to magnetic materials as were specified for dielectrics at the beginning of section 13.6, with the sole exception that hysteresis effects must now be considered at some stage: in this section we ignore them.

Magnetic Energy excluding Non-linear Materials. If all the materials present are para- or diamagnetic, the arguments of section 9.8 are unaffected since the values of L and M include their effects. Thus

$$U_M = \sum_i \tfrac{1}{2}L_i I_i^2 + \sum_i \sum_j M_{ij} I_i I_j = \sum_i \tfrac{1}{2} I_i \Phi_i \qquad (14.25)$$

still. To express U_M in terms of field quantities the argument of section 9.9 can be used except that we must now use the generalized circuital theorem round the tube and replace I by $\oint \mathbf{H}\cdot d\mathbf{s}$ which gives

$$U_M = \iiint \tfrac{1}{2}BH \, d\tau = \iiint B^2/2\mu_r\mu_0 \, d\tau = \iiint \tfrac{1}{2}\mu_r\mu_0 H^2 \, d\tau \quad (14.26)$$

instead of (9.38). (Energy density for anisotropic media is $\frac{1}{2}\mathbf{B}\cdot\mathbf{H}$.)

Magnetic Energy including Non-linear Materials. The argument of section 9.8 must now take account of the fact that the L's and M's are not constant and that $M_{12} \neq M_{21}$ in general. This last fact is exemplified in an ideal transformer consisting of two solenoids of the same length l and cross-section A wound on the same infinite soft ferromagnetic core so that the primary and secondary differ only in the number of turns per unit length (n_p and n_s). A current I in the primary gives $M_{12} = \mu_{rp}\mu_0 n_p n_s l A$ where μ_{rp} is the value of μ_r for $H = n_p I$; while a current I in the secondary gives $M_{21} = \mu_{rs}\mu_0 n_p n_s l A$ where μ_{rs} is for a field $n_s I$. If the core is non-linear, $\mu_{rp} \neq \mu_{rs}$ in general, and $M_{12} \neq M_{21}$.

Taking non-linearity into account, we now find the increment in U_M for small changes of current. For instance, with two circuits carrying currents I_1 and I_2 and with self and mutual inductances L_1, L_2, M_{12} and M_{21}, the increment in U_M when I_1 increases by dI_1 with I_2 constant is $L_1 I_1 \, dI_1 + M_{12} I_2 \, dI_1$ by the same method as in section 9.8. If now I_2 is increased by dI_2 keeping I_1 constant, we find in the limit that

$$dU_M = L_1 I_1 \, dI_1 + L_2 I_2 \, dI_2 + M_{12} I_2 \, dI_1 + M_{21} I_1 \, dI_2$$

and because $\Phi_1 = L_1 I_1 + M_{21} I_2$, etc.

$$dU_M = I_1\, d\Phi_1 + I_2\, d\Phi_2$$

and NOT $\Phi_1\, dI_1 + \Phi_2\, dI_2$ unless $M_{12} = M_{21}$. Thus for any number of circuits,

$$dU_M = \Sigma\, I_i\, d\Phi_i \tag{14.27}$$

which is not equal to $\Sigma\, \Phi_i\, dI_i$ unless all materials are linear. Equation (14.27) can be integrated to give (14.25) only for linear materials.

The argument of section 9.9, applied to obtain U_M in terms of fields, now gives

$$dU_M = \iiint_\tau H\, dB\, d\tau \tag{14.28}$$

so that the increment in energy density is $H\, dB$ (and not $B\, dH$ unless linear). By writing $H\, dB = \mu_0 H\, dH + \mu_0 H\, dM$, it can be seen that if the materials are so weakly magnetic that $\mathbf{H}_{spec}$ is negligible or if the specimen is of such a shape that no poles exist, the first term represents work which would be done in establishing the fields with no material present and the second can therefore properly be associated with the work done in magnetizing the specimen.

Forces between Currents in the Presence of Magnetic Materials. Provided there is no contact between the materials and the current-carrying conductors the forces on the latter will still be given by

$$F_s = (\partial U_M/\partial s)_I = I(\partial\Phi/\partial s) \tag{14.29}$$

as in section 9.11. The forces between currents are thus only likely to be ($\mu_r \times$ the vacuum value) in an infinite LIH medium.

Forces on Linear Magnetic Materials. Like the forces on dielectrics, these calculations are very complex and often of questionable validity unless simplifying assumptions are made which can be reflected in practical situations. Fortunately, the problems of interest refer to weakly magnetic materials in which the fields due to the specimen are so feeble that they may be neglected in comparison with the applied fields. It must be emphasized that many of the formulae to be developed do not apply unless this is so.

We can use either energy methods or force methods as we did with dielectrics. Using the force method first, consider a small volume $d\tau$ of material which has a magnetic moment $\mathbf{M}\, d\tau$ or

$\chi_m \mathbf{H}\, d\tau$. In a non-uniform field it will experience a force given by (7.14) whose x-component is

$$dF_x = \chi_m(H_x\, \partial B_x/\partial x + H_y\, \partial B_x/\partial y + H_z\, \partial B_x/\partial z)\, d\tau$$
$$= \mu_r \mu_0 \chi_m(H_x\, \partial H_x/\partial x + H_y\, \partial H_x/\partial y + H_z\, \partial H_x/\partial z)\, d\tau \qquad (14.30)$$

If no conduction currents flow in the material then $\oint \mathbf{H}\cdot d\mathbf{s} = 0$ and hence by equations similar to (8.34), $\partial H_x/\partial y = \partial H_y/\partial x$ etc., and therefore

$$dF_x = \mu_r \mu_0 \chi_m(H_x\, \partial H_x/\partial x + H_y\, \partial H_y/\partial x + H_z\, \partial H_z/\partial x)\, d\tau \quad (14.31)$$
$$= \tfrac{1}{2}\mu_r \mu_0 \chi_m(\partial H^2/\partial x)\, d\tau = \tfrac{1}{2}\chi_m(\partial B^2/\partial x)\, d\tau/\mu_r \mu_0 \qquad (14.32)$$

If immersed in a fluid medium with a comparable susceptibility the hydrostatic pressures make the calculation very complex. If, however, we can assume that the weakly magnetic material and fluid do not alter $\mathbf{H}$ from its vacuum value by more than a negligible amount, we can assert that were the material to be removed and replaced by the fluid, the force on the same volume of *fluid* would be given by (14.32) with χ_m replaced by χ'_m, this being the susceptibility of the fluid. Since we know that the hydrostatic pressures keep the fluid in equilibrium, we conclude that the forces due to these pressures are opposite to those on the volume $d\tau$ but equal in magnitude. Since these pressures are determined by the fields, they are the same when the material fills the volume and so we finally have

$$dF_x \doteqdot \tfrac{1}{2}\mu_0(\chi_m - \chi'_m)(\partial H^2/\partial x)\, d\tau \qquad (14.33)$$

taking $\mu_r \doteqdot \mu'_r \doteqdot 1$. For a finite volume this must be integrated.

To use the energy method, let U'_M be the energy of the fluid before the magnetic material is inserted. Then with the volume $d\tau$ of material

$$U_M = U'_M + \iiint_\tau \tfrac{1}{2}\mu_r \mu_0 H^2\, d\tau - \iiint_\tau \tfrac{1}{2}\mu'_r \mu_0 H'^2\, d\tau$$

and if we again assume that $\mathbf{H} \doteqdot \mathbf{H}'$, this becomes

$$U_M = U'_M + \iiint_\tau \tfrac{1}{2}(\mu_r - \mu'_r)\mu_0 H^2\, d\tau$$

and hence

$$F_x = \iiint_\tau \tfrac{1}{2}\mu_0(\chi_m - \chi'_m)(\partial H^2/\partial x)\, d\tau \qquad (14.34)$$

since $\mu_r = 1 + \chi_m$. This is the same as (14.33) integrated over τ.

These forces are always such as to cause paramagnetic bodies to move into regions of stronger field irrespective of sign and diamagnetic bodies to move into weaker ones, although a paramagnetic immersed in a fluid of greater positive susceptibility will behave as a diamagnetic. Thus in Fig. 14.8a the diamagnetic will set across the field (this is the usual case considered because the non-uniformity is of the type commonly found between the poles of magnets, but note what happens in Fig. 14.8b). It can be shown

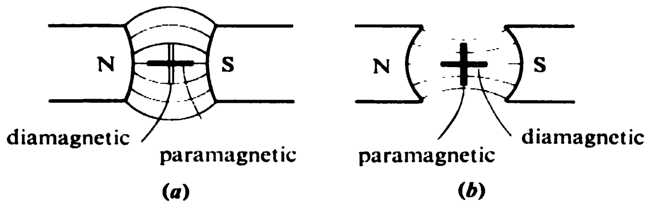


Fig. 14.8. Behaviour of paramagnetic and diamagnetic materials in non-uniform magnetic fields.

that elongated para- or diamagnetic specimens in a *uniform* field experience feeble couples turning them *along* lines of force.

14.6 Ferromagnetic Materials

The properties of ferromagnetic materials are commonly displayed by curves of B against H which may be obtained as explained in section 14.10. In our discussion of the results we shall assume that the specimen is in the form of an anchor ring magnetized by a current in a toroidal coil wound on the ring: since there are no poles, H is simply nI and B is measured by the flux change through a second winding.

If initially unmagnetized, an increase of H gives a B–H curve following OXYZ, the virgin curve (Fig. 14.9). Up to a point X low on the curve a reduction of H causes the curve to be retraced: OX is reversible. Beyond X, a reduction of H causes a path such as that from Y to be followed and, if H is continuously run through a cycle from Y to Y′ and back, the small loop YY′, known as a *hysteresis loop*, is followed. The largest of these loops, ZZ′, is reproducible for a given specimen and is known as *the* hysteresis loop. If at any point such as W′ small reversals of H are made, subsidiary loops like WW′ are traced.

The intercepts on the B-axis denote flux when $H=0$ and thus

indicate the possibility of permanent magnets: OR is called the *remanence*, B_r. The intercepts on the negative H-axis give the reverse fields needed to demagnetize the material, OC being known as the *coercivity*, H_c. We shall see that the so-called demagnetization curve RC contains the useful information about a material intended for use as a permanent magnet.

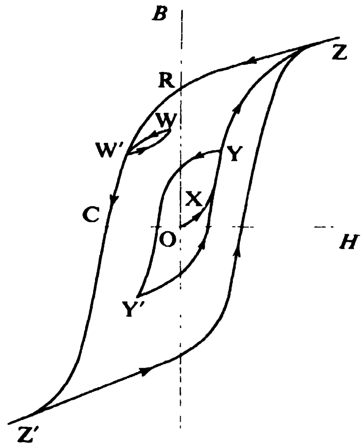


Fig. 14.9. Hysteresis loops.

Incremental Relative Permeability. The value of $\mu_r = B/\mu_0 H$ will vary considerably over the virgin curve, increasing to a maximum and then falling almost to 1. What is often more important is the *differential* or *incremental* μ_r (cf. section 6.2) defined for a small range as $\delta B/\mu_0\,\delta H$. Clearly μ_r and μ_{inc} will only be equal for the initial part of the virgin curve, but when a ferromagnetic forms the core of two coils, one carrying D.C. and the other A.C., the loop WW′ of Fig. 14.9 is typical of that which is followed and μ_{inc} is the slope of WW′ divided by μ_0. The value of μ_{inc} determines the effective inductance and reactance and, since it varies with the operating point on the hysteresis loop and thus with the D.C., the latter can be used to control the impedance of a circuit. Such a device is known as a *saturable reactor.*

Energy Loss in the Hysteresis Loop. We have seen that the energy input needed to increase currents by small amounts can be expressed as $H\,\mathrm{d}B$ per unit volume, even when non-linear materials

are present. If the currents decrease, this amount of energy is recovered so that, without hysteresis, there is no net gain or loss of energy in a complete cycle. Figure 14.10 shows, however, that because the shaded element represents the energy input for a small portion of the loop, the area OAZP gives the energy input for the portion AZ of the loop, while RPZ gives the recovered energy for the portion ZR. Thus OAZR is lost, and when continued round the loop the total energy loss per unit volume per cycle is found to be just the area of the hysteresis loop, which may be written $\oint H \, dB$.

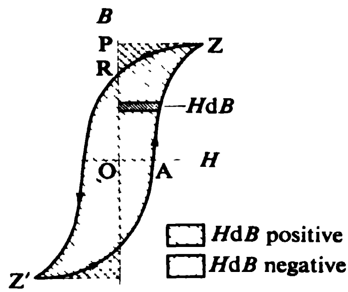


Fig. 14.10. Energy loss in the hysteresis loop.

Magnetization Loop. A graph of M against H shows similar features but becomes horizontal at the extremities due to saturation. (Because $B = \mu_0 H + \mu_0 M$, the value of B is still increasing at saturation, unlike M.) We can also write

$$\oint H \, dB = \oint \mu_0 H \, dH + \oint \mu_0 H \, dM$$

where the left-hand side is the area of the B–H loop, and the right-hand side is the sum of $\mu_0 \times$ the area of the H–H loop (zero) and of $\mu_0 \times$ the area of the M–H loop. Hence the hysteresis loss per unit volume per cycle is also $\mu_0 \times$ the area of the M–H loop.

Soft and Hard Ferromagnetic Materials. In section 14.1, we defined an ideally soft ferromagnetic as one with zero remanence and coercivity. In practice, soft ferromagnetics have a small coercivity, a low loss because of a narrow loop, and a high permeability. Apart from iron itself, used in electromagnets, the material in common use, particularly for large motors, generators and transformers, is silicon–iron (up to 4% Si). Higher permeability and lower loss are possessed by nickel–iron alloys containing

between about 40–80 % Ni (e.g. mumetal, permalloy, supermalloy) but because they generally saturate at lower fields than silicon–iron and are more costly, they are only used in the smaller transformers and chokes: they find extensive other uses, for example as magnetic screens and, because of their large magnetostriction, in transducers.

Hard ferromagnetics used for permanent magnets should have large remanence and coercivity as we see below, but beyond this the properties required depend on the particular use: the largest class of material is that consisting of alloys of Fe, Al, Ni, Co and Cu known generically as Alnico alloys.

Properties of a few representative ferromagnetics are given in table 14.1.

Table 14.1

PROPERTIES OF SOME REPRESENTATIVE FERROMAGNETICS

SOFT	Initial μ_r	Maximum μ_r	H_c (A/m)	H_c (oersted)	B_r (Wb m^2)	B_r (gauss)	$(BH)_{max}$ (J m^3)
4 % Si–Fe	800	8,000	48	0·6			
Mumetal	20,000	80,000	8	0·1			
Supermalloy	100,000	$\sim 10^6$	0·4	0·005			
HARD							
Chromium steel			5,600	70	0·98	9,800	2,270
Alnico (high remanence)			40,000	500	0·80	8,000	13,500
Alcomax III (anisotropic)			52,000	650	1·26	12,600	43,000
Columax (anisotropic)			59,000	740	1·35	13,500	60,000

Data for 4 % Si Fe and mumetal are from de Barr (1953a) by permission of the author; data for hard magnetic materials are from the Permanent Magnet Association's Technical Bulletin No. 1 (1963) by permission. Properties of Supermalloy are given in Boothby and Bozorth (1947).

14.7 The Magnetic Circuit and the Production of Magnetic Fields

Suppose we have a toroidal coil wound closely on an anchor ring formed of sections of various lengths as in Fig. 14.11. If the ring is ferromagnetic there will be little flux leakage, and if it is also thin compared with its diameter both B and H will be approximately constant across any cross-section. The m.m.f. round the ring is

$$\mathscr{H} = \Sigma H_i l_i$$

the summation occurring over the various sections each with their own μ_r, H and in general with differing cross-sections A and flux densities B (the figure does not show this). For each length,

$Hl = Bl/\mu_r\mu_0 = \Phi \times l/\mu_r\mu_0 A$ where Φ is the constant flux in the tube defined by the material. Thus

$$\text{m.m.f.} \qquad \mathscr{H} = \Phi \sum_i \frac{l}{\mu_r\mu_0 A} \tag{14.35}$$

The quantity $l/\mu_r\mu_0 A$, known as the *reluctance* of the length l and denoted by $\mathscr{R}$, is analogous to the resistance of a simple series electric circuit for which

$$\text{e.m.f.} \qquad \mathscr{E} = I \sum \frac{l}{\sigma A} \tag{14.36}$$

σ being the conductivity (cf. equation (6.17)). Because of the analogy between (14.35) and (14.36), it will follow that reluctances

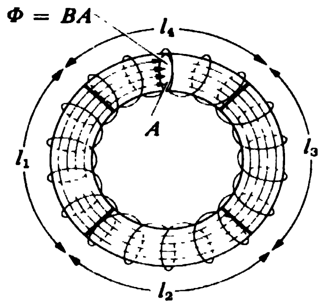


Fig. 14.11. A simple magnetic circuit.

in series and in parallel will combine according to the same laws as resistance.

The equation $\mathscr{H} = \mathscr{R}\Phi$ is not so exact in practice as $\mathscr{E} = RI$ because the assumptions made above do not hold exactly, and because μ_r is not the constant we have assumed. The concept of a magnetic circuit does, however, yield useful results provided any air gaps are small.

Production of Magnetic Fields. The variety in the type of magnetic field required for physical investigation is such that their production is a highly technical matter which can only be touched on here. *Permanent magnets* have the advantage of constancy without power input but cannot produce flux densities even with the best modern materials greater than about 1 Wb/m^2 (10 kgauss

—see table 14.1). *Electromagnets* with iron cores are useful up to about 3 Wb/m^2 and have been constructed to produce as much as 7 Wb/m^2 with the help of over 100 kW input power and a weight of over 35 tons. Beyond 2–3 Wb/m^2, an iron core saturates and any additional flux density is produced by the current while the contribution from the core diminishes: the cost of the core is often not justified by its contribution and for the highest fields *air-cored coils* are used. The great problem here, the dissipation of heat produced by the large currents, is solved by using water-cooling (up to 10 Wb/m^2), superconducting alloys (up to about 5 Wb/m^2) or pulses of current giving transient fields where they can be used (up to 50 Wb/m^2).

We shall look in more detail at electromagnets and permanent magnets using the concept of the magnetic circuit developed above.

Electromagnets. The torus with air gap shown in Fig. 14.12a is a typical electromagnet. If the gap is of length g and the rest of the torus of length l and relative permeability μ_r, the reluctance of the circuit is $l/\mu_r\mu_0 A + g/\mu_0 A$ while the m.m.f. is NI. The flux is therefore

$$\Phi = \frac{\text{m.m.f.}}{\text{reluctance}} = \frac{\mu_0 NIA}{l/\mu_r + g}$$

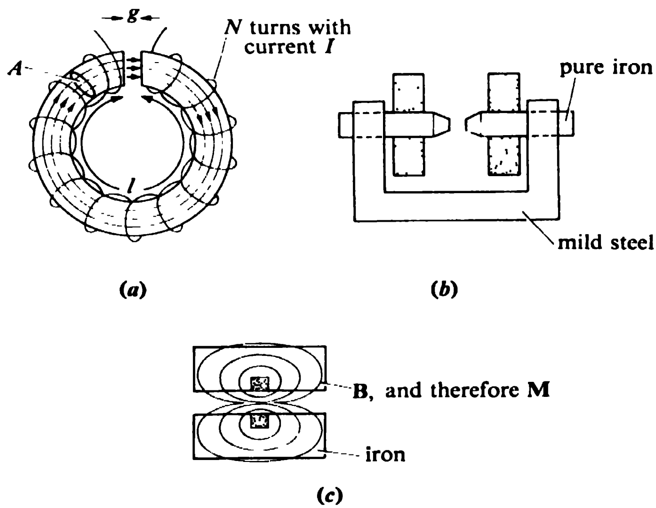


Fig. 14.12. Electromagnets. (a) Magnetic circuit of torus with air gap; (b) Weiss type; (c) Bitter type.

and if the subscript g refers to the gap and m to the material

$$B_m = B_g = \frac{\mu_0 NI}{l/\mu_r + g}; \qquad H_g = \frac{NI}{l/\mu_r + g}; \qquad H_m = \frac{NI}{l + \mu_r g} \tag{14.37}$$

compared with the following values for a complete torus of length $l+g$:

$$B = \mu_r \mu_0 NI/(l+g); \qquad H = NI/(l+g) \tag{14.38}$$

These relations show that for a gap g small compared with l, the opening of the gap in the torus reduces B to $1/(1+\mu_r g/l)$ of its original value, for example to 1% if $g/l = \frac{1}{50}$ and $\mu_r = 5{,}000$. It is also clear that the large reluctance of the gap causes most of the m.m.f. to be concentrated there, giving an H_g larger than H_m.

The Weiss type of electromagnet shown in Fig. 14.12b is very common, and the use of conical pole pieces concentrates the flux over a smaller area, thus increasing B_g and the reluctance of the gap. The Bitter type of magnet (Fig. 14.12c) uses magnetic material more economically than any other type: in it, a single coil, producing a nearly dipolar field, is embedded in a shell of soft ferromagnetic which is magnetized along the lines of force shown. Each magnetized element then produces its maximum field at O (see problem 14.11).

Permanent Magnets. If we could freeze in the magnetization of the gapped torus of Fig. 14.12a while removing the magnetizing coil, we should have the situation of Fig. 14.13 in which the sources and sinks of H (like those of M) occur at the sides of the gap: this is the situation in a permanent magnet. Because the m.m.f. is now zero,

$$H_g g + H_m l = 0 \tag{14.39}$$

and H_m is in the opposite sense to H_g as we should expect: H_m is known as the demagnetizing field. The flux density is continuous, however, and is $\mu_0 H_g$ all round the circuit.

Consider the common form of magnet shown in Fig. 14.13b in which part of the circuit is of soft ferromagnetic of negligible reluctance so that (14.39) still applies but in which the pole pieces define a gap of cross-section a. Then

$$\Phi = B_g a = B_m A$$

and since $H_g g = -H_m l$, the product of the two equations gives

$$B_g H_g \tau_g = -B_m H_m \tau_m$$

where τ_g and τ_m are the volumes of gap and permanent magnet. This equation shows us that to produce a given flux density B_g ($=\mu_0 H_g$) in a given volume τ_g, the smallest volume of material is required when the product $B_m H_m$ is a maximum. Now we have already seen that (14.39) means that H_m is negative if the direction of B is positive so that the state of the ferromagnetic material is given by some point on the demagnetization curve. Figure 14.13c shows this curve together with the product BH plotted against B

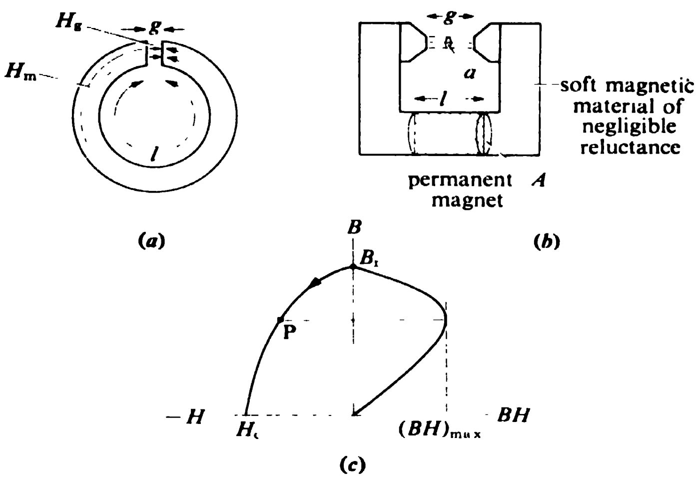


Fig. 14.13. Permanent magnets. (a) Magnetic circuit; (b) use of pole pieces; (c) demagnetization curves.

and we now see that the point P corresponding to $(BH)_{max}$ is the most economical to work at. We also see that for large B_g we require large $(BH)_{max}$ which entails both a high B_r and a high H_c.

A 'keeper' of soft iron placed in the gap has a negligible reluctance and reduces $H_g g$ almost to zero, so that by (14.39), H_m is also reduced almost to zero. This takes the permanent magnet from the working point P on the hysteresis loop back to the point B_r, the demagnetizing field being zero.

The Ideal Permanent Bar Magnet. A true or ideal permanent magnet has a magnetization **M** independent of **H** so that the relations (14.9)–(14.17) apply, but not (14.18)–(14.21). Consider such a magnet as in Fig. 14.14 in the form of a uniformly magnetized bar. The **B** field is due simply to the Amperian currents in the surface

and is similar to that of a finite solenoid. We have seen from (14.17) that sources of **H** are sinks of **M** and vice versa so that the **H** field is as shown, the relation $\mathbf{B}=\mu_0(\mathbf{H}+\mathbf{M})$ applying everywhere Notice that inside the magnet **H** is in opposition to **B** along the axis and that at a general point **B**, **H** and **M** are all in different directions

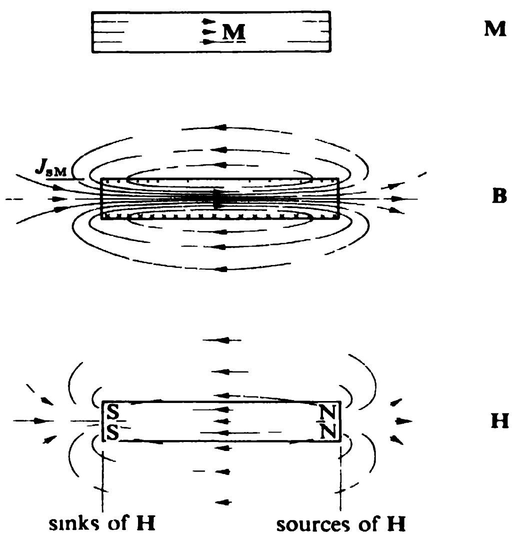


Fig. 14.14. **M**, **B** *and* **H** *in an ideal bar magnet.*

In practice the hysteresis curve shows us that **M** cannot be completely independent of **H** and so the magnetization cannot remain uniform. This is obvious when it is realized that the direction of **H** near the ends would not be parallel to **M** because the sheets of poles have a finite area (cf the electric field in Fig 5 4b) Because of the non-uniformity of **M** thus induced, the **H** field is in turn modified so that a complete solution is very difficult

14.8 Magnetic Poles

We have already discussed the monopole concept and its limitations in chapter 7. For ideal permanent magnets in the absence of conduction currents it is a useful concept because of its similarity to electric charge, and it is possible to develop the theory of magnetization in terms of it. For instance, the dipole moment per unit

volume **M** would be equivalent to a surface and volume distribution of poles, the volume distribution being absent if **M** is uniform or in LIH materials. At a surface, the pole strength per unit area would be equal to the normal component of **M** and thus gives rise to sources of **H** which we should have defined *in vacuo* in chapter 7 by $\mathbf{H}=\mathbf{B}/\mu_0$ so that Gauss's theorem would become $\oint\!\!\oint_S \mathbf{H}\cdot d\mathbf{S}=\Sigma P$. Since *all* poles P are due to magnetization we should have over the surface S, $\Sigma P=-\oint\!\!\oint \mathbf{M}\cdot d\mathbf{S}$ and hence $\oint\!\!\oint(\mathbf{H}+\mathbf{M})\cdot d\mathbf{S}=0$. We should define **B** in a medium by $\mathbf{B}=\mu_0(\mathbf{H}+\mathbf{M})$ so that $\oint\!\!\oint \mathbf{B}\cdot d\mathbf{S}=0$ which is (14.10), while (14.9) would be unaffected since poles are not vortices.

In a bar magnet the essential non-uniformity of **M** in practice, discussed in the previous section, means that there must be a volume distribution of poles near the ends so that only in the limit of a long thin magnet would the resultant pole strength appear to be situated at the ends.

We see that on the pole picture the obvious vector to use is **H**, which has poles as sources. In our methods we can use whichever concept is most useful: the **B** field is given by conduction and Amperian currents as vortices and by no sources, while the **H** field has conduction current as vortices and poles as sources.

The Inverse Square Law between Poles in a Medium. Insofar as single poles cannot be isolated or located with any accuracy it seems fruitless to discuss further a law of force between them. Nevertheless the question is asked as to whether the μ_r occurs in the numerator or the denominator when two point poles are immersed in a medium. All the difficulties arising in the case of Coulomb's law in a medium apply here and in addition there is no way of distinguishing the poles of the medium from those of a magnet since only the discontinuity in **M** gives pole strength. Experimental tests are also impossible because of the absence of a conservation law applying to poles of a single sign and, because no magnet is exactly permanent, the pole strength would always be uncertain.

14.9 Magnetic Fields and Flux Densities inside Materials

The discussion here follows the same lines as in section 13.8 and need not therefore be so extensive. Our model of a magnetic material has been that of a smoothed-out distribution of dipoles which can be shown to be equivalent to surface and volume Amperian currents (**B**-field) or surface and volume poles (**H**-field)

as far as points outside the material and far enough away from it are concerned. Inside materials we have defined **B** and **H** in such a way that they are still the flux density and field strength due to the surface and volume distributions, but we should not expect the *actual* couple per unit dipole or force per unit charge per unit velocity to be equal to **B** since we know that on a microscopic scale the smoothed-out model is not good enough.

As with **E**, there are two cases in which the couple or force would be expected to be **B**. One would think, firstly, that the space and time average of the microscopic **B** would be equal to the macroscopic **B** and that therefore a charge or dipole moving with a high velocity would experience forces and couples appropriately. This has been shown to be so within a few per cent. for mesons and neutrons in iron. The second case is one in which a cavity is scooped out of the material so that the **B**-detecting entity can be placed in it and be 'outside' the material once more. As in the dielectric, there are now surface distributions of Amperian currents or poles produced on the interior surface of the cavity: the reader should have no difficulty in showing that the couple on a dipole in a disc-like cavity perpendicular to **M** is **B**, and in a needle-like cavity along **M** is $\mu_0\mathbf{H}$, while in a spherical cavity it is $\mu_0\mathbf{H}+\frac{1}{3}\mu_0\mathbf{M}$.

14.10 Measurement of Magnetic Permeability and Susceptibility

For substances with susceptibilities differing markedly from 1 it is immaterial whether μ_r or χ_m is the quantity directly measured. For ferromagnetic materials therefore either a B–H or an M–H curve is easily obtained by straightforward electromagnetic methods. For other materials these methods are not sensitive enough and recourse is had to the forces on such materials in non-homogeneous fields, forces which depend directly on χ_m.

Paramagnetic and Diamagnetic Substances. In *Faraday*'s method (also known as *Curie*'s) only a small sample is used and the force given by (14.33), in the form

$$dF_x = \mu_0(\chi_m-\chi_m')\left(H_x\frac{\partial H_x}{\partial x}+H_y\frac{\partial H_y}{\partial x}+H_z\frac{\partial H_z}{\partial x}\right)d\tau$$

is measured. The non-uniformity of **H** is generally such that one term of the three predominates and only dF_x is measured, not dF_y or dF_z. The forces are very small and have led to the development of several delicate balances for their measurement (see Bates, 1961).

In *Gouy*'s method the specimen is in the form of a long rod of

uniform cross-section A perpendicular to the direction x in which the force is to be measured and, by (14.34), since $d\tau = A\,dx$,

$$F_x = \frac{1}{2}\int_0^l \mu_0(\chi_m - \chi'_m)A\frac{\partial H^2}{\partial x}\,dx$$

If the rod has one end in a field H_0 and the other in zero field as in Fig. 14.15,

$$F_x = -\tfrac{1}{2}\mu_0(\chi_m - \chi'_m)H_0^2 A \quad \text{downwards if } \chi_m > \chi'_m$$

and this is generally rather greater than the forces in the Faraday method. For liquids, the specimen is placed in a U-tube and the

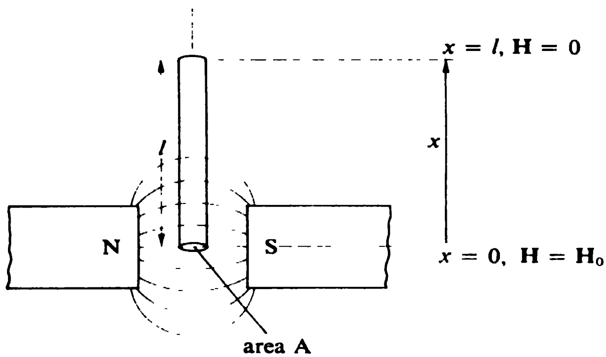


Fig. 14.15. Gouy's method for the measurement of magnetic susceptibility.

force on the liquid in one arm balanced by that due to hydrostatic pressure from the resultant difference in levels (Quincke's method). For gases, χ'_m is determined by using a liquid of known χ_m.

Some typical susceptibilities are shown in table 14.2. Diamagnetic susceptibilities are independent of temperature and their ratio to the density (mass susceptibility see next section) are all of the same order of magnitude. Paramagnetic materials have χ_ρ's differing more widely in magnitude and, except for paramagnetic metals, depending on temperature according to the Curie–Weiss law:

$$\chi_m = \frac{C}{T - \theta_p} \tag{14.40}$$

where T is the absolute temperature, and θ_p a constant for the

material known as its paramagnetic Curie point, which may be positive, negative or zero (in the last case the law is known as Curie's law). Ferromagnetic substances become paramagnetic above a critical temperature θ_f known as the ferromagnetic Curie point and then obey the Curie–Weiss law. θ_f and θ_p are not equal but differ by several degrees for most substances.

Ferromagnetic Substances. A complete hysteresis loop is usually needed here, and the two methods in common use give B and H by the ballistic method, or M and H by the magnetometer method.

The ballistic method measures B in a specimen by the flux change through a coil (the flux coil) wound round it. If the specimen is an anchor ring or torus with no poles, H can be obtained by calculation (nI for a complete toroid): if the specimen is a bar, the ends being connected magnetically by a soft iron yoke to complete the circuit, H is measured by the flux change in a coil (the field-coil) near the surface which links no material (see Figs. 14.16a and b).

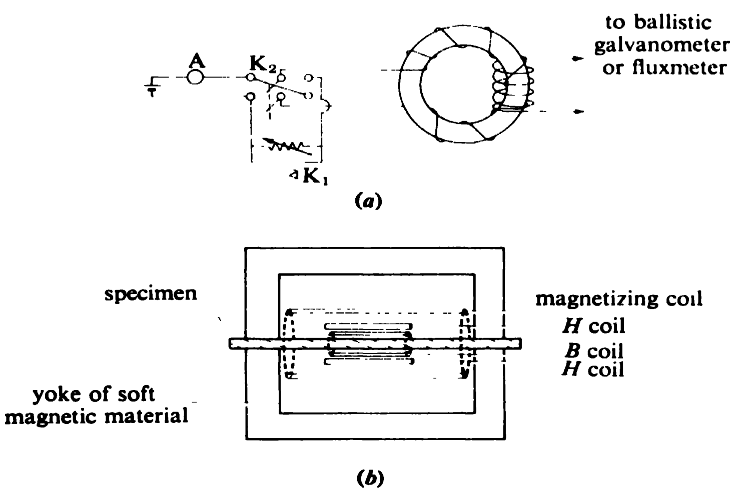


Fig. 14.16. Permeameters with specimens in the form of (a) a torus, (b) a rod.

In the arrangement of Fig. 14.16a the torus is taken to saturation at one end of the hysteresis loop (say Z in Fig. 14.10) and the switch is used either (*a*) to reduce H by opening K_1, thus taking the material to points on the loop between Z and R, (*b*) to reduce and reverse H simultaneously by switch K_2 used with K_1 open or (*c*) to reverse H

to Z′ by switch K_2 with K_1 closed. In this way the loop from Z to Z′ can be plotted, followed by the same procedure from Z′ to Z.

The magnetometer method uses a specimen in the form of a long thin rod whose poles will be located very close to the ends. H is obtained by calculation from the dimensions of the magnetizing solenoid (the demagnetizing field is negligible in a long rod) while M is obtained by treating the rod as a dipole with two poles of strength ($M \times$ area of cross-section) at the ends and measuring the deflection

Table 14.2

SUSCEPTIBILITIES OF SOME DIAMAGNETIC AND PARAMAGNETIC MATERIALS AT 20°C

Diamagnetic	*Susceptibility,** χ_m	*Mass susceptibility,** χ_ρ (m³/kg)
Bismuth	$-16{\cdot}7 \times 10^{-5}$	$-1{\cdot}69 \times 10^{-8}$
Copper	$-0{\cdot}92 \times 10^{-5}$	$-0{\cdot}10 \times 10^{-8}$
Water	$-0{\cdot}91 \times 10^{-5}$	$-0{\cdot}91 \times 10^{-8}$
Hydrogen	$-0{\cdot}22 \times 10^{-8}$	$-2{\cdot}5 \times 10^{-8}$
Paramagnetic		
Aluminium	$2{\cdot}2 \times 10^{-5}$	$0{\cdot}82 \times 10^{-8}$
Platinum	$2{\cdot}6 \times 10^{-5}$	$1{\cdot}22 \times 10^{-8}$
Nitric oxide	$0{\cdot}08 \times 10^{-5}$	$59{\cdot}3 \times 10^{-8}$
Oxygen	$0{\cdot}19 \times 10^{-5}$	133×10^{-8}

* To convert χ_m to unrationalized values (usually called e.m.u.) divide by 4π: to convert χ_ρ to CGS unrationalized values, multiply by $1{,}000/4\pi$.

Data calculated from values given in Kaye and Laby (1959).

on a magnetometer produced by it. The field of the magnetizing solenoid is eliminated at the magnetometer by a compensating coil carrying the same current but situated so that without the rod no deflection is observed for any current.

14.11 The Approach to Microscopic Theory

We introduce two microscopic quantities as in section 13.9: firstly the local magnetic field **F** experienced by a single atom or molecule and secondly the magnetic polarizability α_m defined as the mean magnetic dipole moment per molecule per unit magnetic

field. As in section 13.9 we have

$$\mathbf{M} = \frac{N_A \rho \alpha_m}{M} \mathbf{F}$$

and

$$\chi_m = \frac{N_A \rho \alpha_m}{M} \frac{\mathbf{F}}{\mathbf{H}} \tag{14.41}$$

The mass susceptibility, $\chi_\rho = \chi_m/\rho$, and the molar susceptibility, $\chi_A = \chi_m M/\rho$, are also used so that, for instance, $\chi_A = N_A \alpha_m \mathbf{F}/\mathbf{H}$.

Local Field. As in a dielectric, the couple exerted on a molecular dipole will not be that due just to the macroscopic $\mathbf{B}$ except in gases at low pressure: in general the couple will be $\mu_0\mathbf{H} + \mu_0 N\mathbf{M}$ where $N\mathbf{M}$ is the *molecular field.* The Lorentz method would give $N = \frac{1}{3}$ as in the case of dielectrics but in general we have from (14.41) that

$$\chi_m = \frac{K}{1-KN} \quad \text{where} \quad K = N_A \rho \mu_0 \alpha_m / M \tag{14.42}$$

Magnetic Carriers. We now ask how atoms and molecules give rise to the observed magnetic polarizability α_m if every type of fundamental magnetic dipole has a magnetic moment $\mathbf{m}$ and an angular momentum $\mathbf{L}$ related by $\mathbf{m} = \gamma \mathbf{L}$ (section 11.5).

Two gyromagnetic experiments show that $\mathbf{m}$ and $\mathbf{L}$ are always associated and allow γ to be determined. The Barnett effect is the magnetization of a rod by rotation causing the alignment of angular momenta. The Einstein–de Haas effect is the change $\delta\mathbf{L}$ in internal angular momentum which accompanies a change in magnetization $\delta\mathbf{M}$: in this effect the conservation of angular momentum ensures that there will be an equal and opposite change in the angular momentum of the atomic arrangement as a whole which reveals itself macroscopically. Both effects show that γ is negative, but that the g-factor (section 11.5) is not always 1 as it would be if due to orbital motion only. The concept of electron spin with an associated γ almost exactly e/m and $g \doteq 2$ has to be invoked to explain optical spectra, and gyromagnetic experiments with ferromagnetics give g's very nearly equal to 2, suggesting that electron spin is almost entirely responsible for it.

Diamagnetism. If $\mathbf{m} = 0$ for an atom or ion (usually because the electronic structure consists of completed shells) the only magnetic effect occurring is the precession of the orbits discussed in section 11.5. The sense of the precession is such as to produce a flux opposing the applied field and thus accounts for μ_r being less than 1.

The effect is internal to each atom and is thus independent of temperature. Diamagnetism is a universal effect but, like the distortion term in a dielectric, is usually swamped by the effect of a permanent **m** when present.

Paramagnetism. An intrinsic moment is possessed by an atom or ion when one or more of its electron shells are incomplete, and it can only be calculated by quantum mechanical methods. The contribution of the atomic moment **m** to the macroscopic moment **M** depends on its alignment with respect to the magnetic field which is again determined by quantum considerations. Changes of alignment may take place through interactions (e.g. collisions) between one atom and another and paramagnetism is thus dependent on temperature. A simple theory shows that α_m for moderate fields is given by $\mathbf{m}^2/3kT$ where k is Boltzmann's constant and T the absolute temperature, and substitution of this in (14.42) yields the Curie-Weiss law (the Curie law if $N = 0$). To sum up, while we can explain a μ_r greater than 1 qualitatively by a partial alignment of the **m**'s, a detailed explanation must be a quantum one.

Ferromagnetism. There are two separate aspects of ferromagnetism. The first is the large magnetization which can occur in the absence of an external field: this is due to the spontaneous magnetization caused by an alignment of electron spins which once more requires quantum mechanics to provide even a partial explanation. The second is the fact that the spontaneous magnetization can apparently be destroyed and recreated in a large specimen with accompanying hysteresis. This is known to be due to the existence of small *domains* of the material each spontaneously magnetized to saturation in a definite crystalline direction, adjacent domains having different directions. These domains (Fig. 14.17), of size about 20 μ, are separated by walls about 50 Å thick in which the direction of the spins is changing. Domains are formed in the first place even in a single crystal because the energy is lowered thereby, but although smaller domains mean less energy the wall area increases and the process ceases when the decrease in energy due to a splitting into smaller domains equals the increase due to the resultant increase in wall area.

The application of an increasing magnetic field causes (*a*) initially, reversible wall movements allowing domains more favourably magnetized to grow at the expense of others, (*b*) for moderate fields, irreversible wall movements of the same type but hindered by impurities or defects in the crystal structure and (*c*) for high fields,

rotation of magnetization from easy directions towards that of the field. Each of these stages accounts for part of the hysteresis loop and the irreversible step (*b*) can be detected audibly by e.m.f.s induced in loudspeaker coils wound on a specimen as a wall jumps irreversibly across a defect (the Barkhausen effect). The patterns

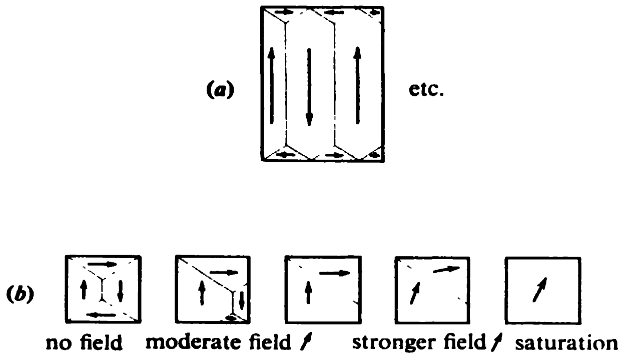


Fig. 14.17. Ferromagnetic domains.

of domains on the surfaces of materials can be made visible under a microscope by fine ferromagnetic powders and the wall movements can be watched as a magnetic field is applied.

14.12 Summary of Chapter 14

By introducing the magnetic field strength **H** we have been able to summarize the magnetic laws by

$$\oint\!\!\oint \mathbf{B}\cdot d\mathbf{S} = 0 \quad \text{and} \quad \oint \mathbf{H}\cdot d\mathbf{s} = I_c \quad (14.11) \text{ and } (14.15)$$

where $\mathbf{B}=\mu_0(\mathbf{H}+\mathbf{M})$ and **M** is defined as the magnetic dipole moment per unit volume of the material. Thus **B** has no sources, although both conduction and Amperian currents are vortices; while **H** has conduction currents only as vortices and sinks of **M** (poles) as sources.

Magnetic materials are described by the vectors **M** and **H** between which the relation

$$\mathbf{M} = \chi_m \mathbf{H} \qquad (14.18)$$

exists, χ_m being the magnetic susceptibility, a scalar constant only in LIH materials: in ferromagnetics it is not even single-valued and

depends on the previous history. In ideally hard ferromagnetics **M** is a constant independent of **H** and a χ_m does not exist.

If we define

$$\mu_r = 1+\chi_m \tag{14.21}$$

in general, μ_r has the same properties as those of the relative permeability defined as L_m/L_0. It follows that, also in general,

$$\mathbf{B} = \mu_r\mu_0\mathbf{H} \tag{14.22}$$

and that the boundary conditions on **B** and **H** are that the normal component of **B** is continuous and the tangential component of **H** is discontinuous by the conduction surface current density.

We have not considered time-varying fields: these we leave until the next chapter.

Appendix 14.1 Definitions of H and χ_m

The substitution $\mu_0 \to 4\pi$ used in chapters 7–9 to convert formulae to CGS e.m.u. form does not work for relations involving **H** and χ_m. The reason for this is that the definitions differ by a factor 4π:

$$\mathbf{H}_{\text{MKSA}} = \mathbf{B}/\mu_0 - \mathbf{M} \quad \text{whereas} \quad \mathbf{H}_{\text{CGS}} = 4\pi(\mathbf{B}/\mu_0 - \mathbf{M})$$

$$\chi_{\text{mMKSA}} = \mathbf{M}/\mathbf{H} \quad \text{whereas} \quad \chi_{\text{mCGS}} = \mathbf{M}/4\pi\mathbf{H}$$

Hence the substitutions needed to convert all formula in this chapter to CGS e.m.u. are

$$\mu_0 \to 4\pi; \qquad \mathbf{H} \to \mathbf{H}/4\pi; \qquad \chi_m \to 4\pi\chi_m \tag{14.43}$$

so that for instance

$$\mathbf{B} = \mathbf{H}+4\pi\mathbf{M}; \qquad \mathbf{M} = \chi_m\mathbf{H}; \qquad \mu_r = 1+4\pi\chi_m \quad \text{CGS e.m.u.} \tag{14.44}$$

Appendix 14.2 Sommerfeld and Kennelly Magnetization

Because **m** is defined differently in the S and K versions of the MKSA system (by a factor μ_0—see appendix 7.3), the magnetization is also different. We have used the S version throughout the chapter (though British and the older American MKS texts prefer the K version—see section 16.7), but we summarize here the alternative.

The K version defines the magnetic moment per unit volume, **I**, in Wb/m^2 so that in the derivation of the surface density of Amperian current (14.8), the moment of the element in Fig. 14.3 is $\mu_0 J_s \, dS \, d\mathbf{l}$

and $\mathbf{J}_s$ is therefore $\mathbf{I}/\mu_0$, and this must replace $\mathbf{M}$ throughout. Thus in the derivations of section 14.3, the circuital theorem (14.11) becomes $\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0 I_c + \mu_0 I_M$ in which $I_M = \mathbf{I}\cdot d\mathbf{s}/\mu_0$ and hence (14.12) becomes $\oint (\mathbf{B}/\mu_0 - \mathbf{I}/\mu_0)\cdot d\mathbf{s} = I_c$.

$\mathbf{H}$ is now defined as

$$\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{I}/\mu_0 \quad \text{or} \quad \mathbf{B} = \mu_0\mathbf{H} + \mathbf{I}$$

so that the generalized circuital theorem (14.15) follows.

The magnetic susceptibility χ_m is defined by

$$\mathbf{I} = \mu_0 \chi_m \mathbf{H}$$

The K version follows more closely the dielectric one and is thus more appropriate when the pole model is used. To convert Kennelly formulae to CGS e.m.u., the substitution $\mathbf{I} \rightarrow 4\pi\mathbf{I}$ must be made in addition to (14.43) so that, for instance,

$$\left.\begin{aligned} \mathbf{B} &= \mathbf{H} + 4\pi\mathbf{I} \\ \mathbf{I} &= \chi_m \mathbf{H} \\ \mu_r &= 1 + 4\pi\chi_m \end{aligned}\right\} \text{CGS e.m.u.}$$

The conversion from the S version in this chapter to the K version is achieved by the substitution $\mathbf{M} \rightarrow \mathbf{I}/\mu_0$.

$\mathbf{I}$ is sometimes called the *magnetic polarization* to distinguish it from $\mathbf{M}$.

References

Experimental methods: Bates (1961). Theories of diamagnetism, paramagnetism, ferromagnetism: Cusack (1958). Ferromagnetic materials: de Barr (1953a, 1953b); Brailsford (1960). Production of magnetic fields: Bitter (1960); Olsen (1964); Montgomery (1963).

PROBLEMS

SECTION 14.1

14.1 A ferromagnetic rod of relative permeability 10^4 is pushed into a long solenoid. If the rod fills the whole of space in which any appreciable magnetic flux density occurs, find its effect on (*a*) the time constant of the solenoid, (*b*) the resonant frequency when connected to a condenser and (*c*) the Q-factor of the circuit in (*b*), assuming no losses in the rod.

14.2 A long solenoid with 15 turns per cm each carrying 0·1 A is wound on an even longer iron core ($\mu_r = 1{,}000$). Compare the magnitudes of

conduction and Amperian surface current densities. What does the Amperian surface current density become if the rod is of copper ($\mu_r \approx 1 - 10^{-5}$)?

SECTION 14.2

14.3 An anchor ring of cross-section 2 cm^2, mean radius 20 cm and $\mu_r = 1{,}500$ is closely and uniformly wound with 2,000 turns of wire. Calculate the self-inductance of the toroidal coil. If a current of 0·1 A is passed through the wire, find the mean magnetic flux density and the magnetization in the anchor ring.

*14.4 Show that, provided $\mathbf{M} = \sigma\boldsymbol{\omega}a$, identical magnetic fields are produced by (*a*) a ferromagnetic sphere with uniform magnetization **M** and (*b*) a non-ferromagnetic conducting sphere carrying a uniform surface density of charge σ and rotating about a diameter with angular velocity $\boldsymbol{\omega}$, a being the radius of both spheres.

SECTION 14.3

14.5 Find the mean magnetic field strength in the anchor ring of problem 14.3.

SECTION 14.4

14.6 The inner conductor of a coaxial cable has a radius a and is coated with a sleeve of non-conducting ferromagnetic whose external radius is $2a$. If the inner radius of the outer conductor is $4a$ and the relative permeability of the ferromagnetic is μ_r, find the self-inductance per unit length of the cable neglecting the flux in the material of the conductors.

14.7 Derive the condition governing the refraction of the lines of **B** and **H** across a plane boundary between two LIH media. Explain why such refraction is in practice extremely small.

SECTION 14.7

14.8 Show that the increment in magnetic energy per unit volume in the magnetic circuit of Fig. 14.11 may be written as $\mathscr{H}\, d\Phi$.

14.9 What is the magnetic flux density in a gap of width 5 mm opened in the anchor ring of problem 14.3?

14.10 Find the flux through the centre arm of the magnetic circuit shown in Fig. 14.18 if the cross-section of all arms is 20 cm^2, if the relative permeability is 1,000 and the magnetizing coil is of 1,000 turns carrying 0·2 A.

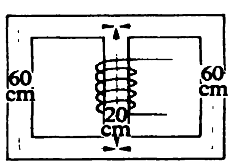


Fig. 14.18. Magnetic circuit for problem 14.10.

*14.11 Show that the magnetic flux density at the origin produced by a small magnetic dipole situated at the point (r, θ) and making an angle α with r is a maximum when $\tan\alpha = \frac{1}{2}\tan\theta$. Hence explain the principle of the Bitter magnet of Fig. 14.12c.

SECTION 14.8

14.12 Show that, inside a uniformly magnetized permanent magnet, **B** and **H** are always in opposite directions.

CHAPTER 15

MAXWELL'S EQUATIONS AND ELECTROMAGNETIC WAVES

Our study of electricity and magnetism has now reached the point at which we must take stock of the laws which have been established in previous chapters and see whether they need any further generalization. It is the merit of using field quantities (**E**, **B**, **D** and **H**) to express the laws that they reveal an inconsistency only resolved by the addition of the so-called displacement current, an addition first suggested by Maxwell which completes the set of equations known by his name and which has far-reaching consequences, some of which are examined in the later sections of this chapter. Complete generalization is beyond the scope of this volume and we restrict the treatment now to media which are stationary, linear, isotropic and homogeneous.

15.1 Summary of Electric and Magnetic Laws

The general forms of the laws so far established are expressed in terms of the electric field strength **E** and the magnetic flux density **B** which we now take to be defined by the Lorentz force formula

$$\mathbf{F} = Q(\mathbf{E}+\mathbf{v}\times\mathbf{B}) \qquad (11.3)=(15.1)$$

or, if we prefer to accept that the charge Q is small but finite in size

$$\mathrm{d}\mathbf{F} = (\rho\mathbf{E}+\mathbf{J}\times\mathbf{B})\,\mathrm{d}\tau \qquad (15.2)$$

since $\mathbf{J}=\rho\mathbf{v}$.

Electrostatics, generalized to take account of dielectrics, is summarized by

$$\oiint \mathbf{D}\cdot\mathrm{d}\mathbf{S} = \Sigma\, Q \quad \text{or} \quad \iiint \rho\,\mathrm{d}\tau \qquad (13.10)=(15.3)$$

$$\oint \mathbf{E}\cdot\mathrm{d}\mathbf{s} = 0 \qquad (13.9)=(15.4)$$

both of which follow from Coulomb's law and the definition of **D**.

The magnetic effects of steady currents in the presence of magnetic media are summarized by

$$\oiint \mathbf{B}\cdot d\mathbf{S} = 0 \qquad (14.11)=(15.5)$$

$$\oint \mathbf{H}\cdot d\mathbf{s} = \Sigma I_c \quad \text{or} \quad \iint \mathbf{J}_c\cdot d\mathbf{S} \qquad (14.15)=(15.6)$$

which follow from the absence of monopoles, the properties of magnetic dipoles and the definition of **H**. The Q and I_c are respectively conduction charge and current only.

These last four equations apply to steady fields and table 15.1 summarizes them in terms of the sources and vortices we have previously used as a help in picturing what the equations represent.

Table 15.1

STEADY FIELDS

	Sources	*Vortices*
D	Conduction charge	(Vortices of **P**)
E	All charge	None
B	None	All current
H	(Sinks of **M** = poles)	Conduction current

It is possible to draw parallels between **E** and **B** on the one hand and between **D** and **H** on the other and this is probably the natural way to look at the four when the ultimate origin of all magnetism is assumed to be circulating currents. It is also possible to draw parallels between **E** and **H** on the one hand and **D** and **B** on the other, but we have seen that this is more appropriate to a magnetism based on poles. The view is taken here that such associations are no more than aids to understanding and that no special significance should be attached to them.

Time Variations. In chapter 9 the effect of varying a magnetic field (with stationary media, which we are assuming) was seen to result in Neumann's law $\mathscr{E} = -\partial\Phi/\partial t$. Writing $\mathscr{E}$ as $\oint_C \mathbf{E}\cdot d\mathbf{s}$ and Φ as $\iint_S \mathbf{B}\cdot d\mathbf{S}$, where S is the surface bounded by the path C,

$$\oint_C \mathbf{E}\cdot d\mathbf{s} = -\frac{\partial}{\partial t}\iint_S \mathbf{B}\cdot d\mathbf{S} = \iint_S -\frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{S} \qquad (15.7)$$

which means that the vortices of **E** and **D** in table 15.1 should now include time-varying **B**.

Equation (15.7) does not contradict (15.4) but in fact includes it as a special case when $\partial \mathbf{B}/\partial t = 0$ everywhere. The question now to be resolved is whether any of the other three equations which were derived for steady fields have to be modified when time variations occur.

15.2 Displacement Current

Consider a parallel-plate condenser as in Fig. 15.1 being charged by a conduction current I_c. The magnetic field strength **H** round a path C satisfies (15.6), i.e. $\oint_C \mathbf{H}\cdot d\mathbf{s} = I_c$ where I_c is the current encircled by the path C. If we ask the question 'how do we decide what encirclement means?', we should reply in the case of a steady current that *any* surface S bounded by C should be taken and the current across it found. In Fig. 15.1, however, while S_1 is certainly

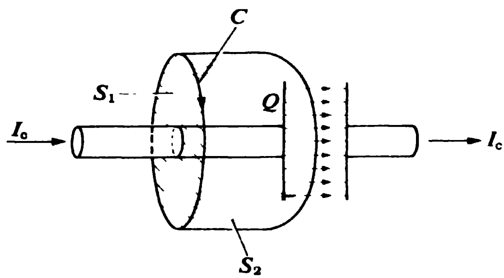


Fig. 15.1. Application of the circuital theorem to a changing conduction current.

crossed by I_c, S_2 has no conduction current across it and so we have both

$$\begin{aligned} A&: \quad \oint_C \mathbf{H}\cdot d\mathbf{s} = I_c \quad \text{if } S_1 \\ B&: \quad \oint_C \mathbf{H}\cdot d\mathbf{s} = 0 \quad \text{if } S_2 \end{aligned} \tag{15.8}$$

We are in no doubt that at some distance from the condenser the correct expression is A since we can measure **H** and confirm it; and we are equally sure that **H** does not satisfy alternative B. There is here an inconsistency: some way must therefore be found of generalizing the circuital theorem so that the correct expression A is obtained independent of which surface S is chosen.

Some indication of the solution is given by the special case of Fig. 15.1. While S_2 has no flux of conduction current density J_c across it, there is a flux of **D** *which changes as long as I_c flows.* Moreover, the rate of change of this flux is equal to I_c for, since $Q = DA$ where A is the area of the plates, $I_c = \mathrm{d}Q/\mathrm{d}t = \mathrm{d}(DA)/\mathrm{d}t$. If now the right-hand side of the circuital theorem is amended to $I_c + I_d$ where

$$I_d = \mathrm{d}(DA)/\mathrm{d}t \tag{15.9}$$

we should have that for S_2, $I_c = 0$, and for S_1, $I_d = 0$, so that $\oint_c \mathbf{H}\cdot\mathrm{d}\mathbf{s}$ would be the same for any S bounded by C.

I_d is known as the *displacement current*, and across any surface should be given by the rate of change of the flux of **D**:

$$I_d = \frac{\partial}{\partial t}\iint_S \mathbf{D}\cdot\mathrm{d}\mathbf{S} = \iint_S \frac{\partial \mathbf{D}}{\partial t}\cdot\mathrm{d}\mathbf{S} \quad \text{(Definition of } I_d) \tag{15.10}$$

By analogy with conduction current, $\partial\mathbf{D}/\partial t$ can be regarded as a displacement current density $\mathbf{J}_d$.

Generalization. In general, both S_1 and S_2 could have conduction and displacement currents across them and the situation of Fig. 15.2 arises. Here, the current entering the volume bounded by S_1

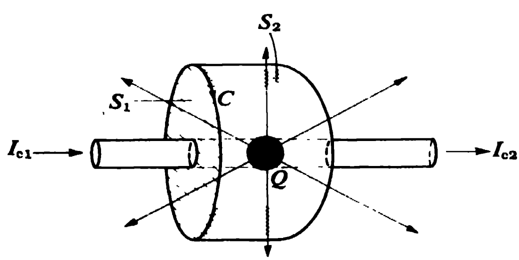


Fig. 15.2. Generalization of displacement current.

and S_2 is greater than that leaving, so that the charge Q within the volume increases according to

$$I_{c1} - I_{c2} = \mathrm{d}Q/\mathrm{d}t \tag{15.11}$$

Applying Gauss's theorem to the same volume gives

$$\iint_{S1} \mathbf{D}\cdot d\mathbf{S} + \iint_{S2} \mathbf{D}\cdot d\mathbf{S} = Q$$

and taking the time rate of change of both sides

$$-I_{d1} + I_{d2} = dQ/dt \tag{15.12}$$

the negative sign occurring because I_d must be related to the sense of C by the right-hand screw rule. (15.11) and (15.12) give

$$I_{c1} + I_{d1} = I_{c2} + I_{d2}$$

and the sum of conduction and displacement current is independent of S. It is therefore postulated that the generalization of (15.6) is

$$\oint \mathbf{H}\cdot d\mathbf{s} = I_c + I_d \tag{15.13}$$

The additional term implies that a magnetic field **H** accompanies a changing electric flux just as **E** accompanies a changing magnetic flux: it would also add time-varying electric displacement to the vortices of **H** and **B** in table 15.1.

If (15.13) is correct why have we been able to obtain verifiable results without the displacement current term? The answer for steady currents is that, referring to Fig. 15.2, $I_{c1} = I_{c2}$ so that Q and **D** are constant and the displacement currents are zero. In A.C. circuits we confined our attention to the connecting wires so that, referring to Fig. 15.1, only I_c was taken into account: this will give correct results as long as the electric field is confined to the region between the condenser plates (for then $I_c = I_d$ and we have seen that *either* may be used) and also as long as the variations in current are slow enough for I_c to be the same at all points in the connecting wires (the assumption of quasi-steady conditions). We therefore expect any effects due to displacement current to become noticeable when high frequency electric fields occur.

Problem 15.2 reveals one case in which only displacement current can give a solution: the magnetic flux density of a charge moving with a uniform velocity previously met in (8.30). Here the application of the old circuital theorem (15.6) would give $\mathbf{B} = \mu_0\mathbf{H} = 0$ and thus predict no magnetic effect from a single moving charge and therefore presumably from a current. (See also French and Tessman, 1923.)

15.3 Maxwell's Equations

In Integral Form. The complete set of electromagnetic laws now reads:

$$\oiint_S \mathbf{D}\cdot d\mathbf{S} = \Sigma Q \quad \text{or} \quad \iiint_\tau \rho \, d\tau$$

Coulomb's law; definition of **D** (15.14)

$$\oint_C \mathbf{E}\cdot d\mathbf{s} = -\iint_S \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{S}$$

Coulomb's law; Neumann's law (15.15)

$$\oiint_S \mathbf{B}\cdot d\mathbf{S} = 0$$

Absence of monopoles; circuital theorem (15.16)

$$\oint_C \mathbf{H}\cdot d\mathbf{s} = I_c + \iint_S \frac{\partial \mathbf{D}}{\partial t}\cdot d\mathbf{S} \quad \text{or} \quad \iint_S \mathbf{J}_c\cdot d\mathbf{S} + \iint_S \frac{\partial \mathbf{D}}{dt}\cdot d\mathbf{S}$$

Circuital theorem; definition of **H**; displacement current (15.17)

in which the surfaces S are bounded by the paths C, the volume τ by the surface S, and Q and I include conduction charge and current only. These are Maxwell's equations in integral form and we adopt the hypothesis that they need no further generalization.

In Differential Form (a) in Cartesians. It is often more useful to express Maxwell's equations in a differential form giving the relations between the quantities at a point. The method of section 4.2 applied to the surface and line integrals in (15.14)–(15.17) yield the following:

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho \quad \text{from (15.14)} \tag{15.18}$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \quad \text{from (15.16)} \tag{15.19}$$

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t}; \qquad \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{\partial B_y}{\partial t};$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{\partial B_z}{\partial t} \quad \text{from (15.15)} \tag{15.20}$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = J_x + \frac{\partial D_x}{\partial t}; \qquad \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = J_y + \frac{\partial D_y}{\partial t};$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = J_z + \frac{\partial D_z}{\partial t} \quad \text{from (15.17)} \tag{15.21}$$

In Differential Form (b) Using Vector Operators. For the benefit of those who are familiar with vector field theory, and to encourage others to become familiar with it, the following are the forms taken by (15.18) to (15.21) using the vector operators div (=divergence) and curl (see problem 8.16):

$$\left.\begin{aligned} &\nabla\cdot\mathbf{D} \equiv \operatorname{div}\mathbf{D} = \rho; \qquad \nabla\cdot\mathbf{B} \equiv \operatorname{div}\mathbf{B} = 0; \\ &\nabla\times\mathbf{E} \equiv \operatorname{curl}\mathbf{E} = -\partial\mathbf{B}/\partial t; \qquad \nabla\times\mathbf{H} \equiv \operatorname{curl}\mathbf{H} = \mathbf{J}+\partial\mathbf{D}/\partial t \end{aligned}\right\} \tag{15.22}$$

where ∇ in cartesians stands for $\mathbf{i}\,\partial/\partial x+\mathbf{j}\,\partial/\partial y+\mathbf{k}\,\partial/\partial z$ (cf. the Laplacian ∇^2 in section 4.2). The conciseness of (15.22) and the independence of any particular co-ordinate system (since the operators div and curl can be obtained in any system when required) make it a desirable form for any extensive development of electromagnetic theory beyond this stage. We shall rely on the cartesian form of the equations with an occasional indication of the method to be adopted using vector operators.

15.4 Electromagnetic Waves *in Vacuo*

The previous chapters of this book have dealt in detail with applications of Maxwell's equations in the absence of displacement current so that it is the effect of this extra term which is of greatest interest now. We shall look for solutions giving the fields *in vacuo* and in infinite LIH non-ferromagnetic media (section 15.5) for which the relations between **D** and **E** and between **B** and **H** take the particularly simple forms $\mathbf{D}=\varepsilon_r\varepsilon_0\mathbf{E}$ and $\mathbf{H}=\mathbf{B}/\mu_r\mu_0$: to go beyond this would take us outside the scope of this volume.

In vacuo, the region in which we are interested contains no conduction currents or charges and, since $\mathbf{D}=\varepsilon_0\mathbf{E}$ and $\mathbf{H}=\mathbf{B}/\mu_0$, Maxwell's equations become

$$\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z} = 0 \qquad \frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}+\frac{\partial B_z}{\partial z} = 0$$

$$\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t} \qquad \frac{\partial B_z}{\partial y}-\frac{\partial B_y}{\partial z} = \mu_0\varepsilon_0\frac{\partial E_x}{\partial t}$$

$$\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x} = -\frac{\partial B_y}{\partial t} \qquad \frac{\partial B_x}{\partial z}-\frac{\partial B_z}{\partial x} = \mu_0\varepsilon_0\frac{\partial E_y}{\partial t}$$

$$\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y} = -\frac{\partial B_z}{\partial t} \qquad \frac{\partial B_y}{\partial x}-\frac{\partial B_x}{\partial y} = \mu_0\varepsilon_0\frac{\partial E_z}{\partial t}$$

Let us pick one of the field components, say E_y, and find an equation for it by eliminating the other components of **E** and all those of **B**. The steps are not difficult to follow using the above equations.

$$
\begin{aligned}
\mu_0\varepsilon_0 \frac{\partial^2 E_y}{\partial t^2} = \frac{\partial}{\partial t}\left(\mu_0\varepsilon_0 \frac{\partial E_y}{\partial t}\right) &= \frac{\partial}{\partial t}\left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x}\right) \\
&= \frac{\partial}{\partial z}\left(\frac{\partial B_x}{\partial t}\right) - \frac{\partial}{\partial x}\left(\frac{\partial B_z}{\partial t}\right) \\
&= \frac{\partial}{\partial z}\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y}\right) - \frac{\partial}{\partial x}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x}\right) \\
&= \frac{\partial^2 E_y}{\partial z^2} + \frac{\partial^2 E_y}{\partial x^2} - \frac{\partial}{\partial z}\left(\frac{\partial E_z}{\partial y}\right) - \frac{\partial}{\partial x}\left(\frac{\partial E_x}{\partial y}\right) \\
&= \frac{\partial^2 E_y}{\partial z^2} + \frac{\partial^2 E_y}{\partial x^2} - \frac{\partial}{\partial y}\left(\frac{\partial E_z}{\partial z} + \frac{\partial E_x}{\partial x}\right) \\
&= \frac{\partial^2 E_y}{\partial z^2} + \frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial y^2}
\end{aligned}
$$

or
$$\nabla^2 E_y - \mu_0\varepsilon_0 \frac{\partial^2 E_y}{\partial t^2} = 0 \tag{15.23}$$

and it is clear that the other components of **E** and those of **B** would obey the same equation. (15.23) is the equation obeyed by waves travelling with a velocity $1/(\mu_0\varepsilon_0)^{1/2}$, an equation which may be more familiar to readers in its one-dimensional form as it would apply to waves on a string for instance:

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} = 0$$

where ϕ is the displacement. The velocity of electromagnetic waves *in vacuo* will be denoted by c.

In terms of vector operators, using (15.22), we take the curl of both sides of the third equation and use the identity (curl curl) = (grad div $-\nabla^2$):

$$\text{curl curl } \mathbf{E} = -\frac{\partial}{\partial t}(\text{curl } \mathbf{B})$$

or $\quad \text{grad div } \mathbf{E} - \nabla^2\mathbf{E} = -\mu_0\varepsilon_0 \dfrac{\partial^2 \mathbf{E}}{\partial t^2} \quad$ since $\quad \text{curl } \mathbf{B} = -\partial\mathbf{E}/\partial t$

or $$\nabla^2 \mathbf{E} - \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \qquad \text{since div } \mathbf{E} = 0$$

of which (15.23) is the y-component.

In the MKSA system of units used here, μ_0 is assigned the value $4\pi \times 10^{-7}$ and ε_0 has a value which can be determined by electrical experiments (sections 2.4, 5.4 and 16.7) so that the velocity of electromagnetic waves may be obtained from purely electrical measurements: its value is almost exactly 3×10^8 m/s. It was the great achievement of Maxwell not only to predict the existence of such waves but to suggest that light, which it was known travelled with this same velocity, was a form of electromagnetic wave. Heinrich Hertz was among the first to demonstrate the existence of waves generated by the oscillatory spark discharge and to show that they possessed many of the familiar properties of light such as reflection, refraction, interference and polarization. It is now known that a complete spectrum of such waves exists with wave-lengths ranging from the several metres of the long radio waves, through the shorter radio waves and microwaves, the infra-red, visible and ultra-violet to the X-rays with wave-lengths of only a few ångstrom units (and comparable with interatomic distances) and beyond these to γ-rays (Fig. 15.3).

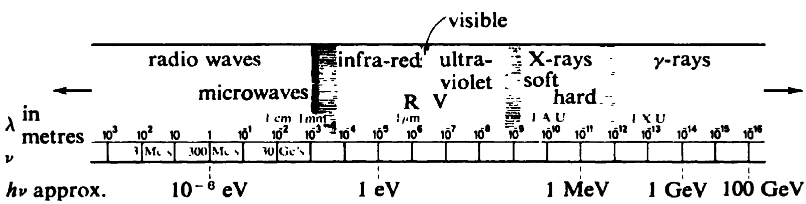


Fig. 15.3. The electromagnetic spectrum.

Plane Waves. By a plane wave is meant one in which the values of all field quantities are constant over a plane perpendicular to the direction of propagation. If we take the x-axis as the direction of propagation, the plane wave-fronts are parallel to the yz-plane and $\partial/\partial y$ and $\partial/\partial z$ of any quantity are zero. Maxwell's equations thus reduce to

$$\partial E_x/\partial x = 0; \qquad \partial B_x/\partial x = 0; \qquad \partial B_x/\partial t = 0; \qquad \partial E_x/\partial t = 0 \tag{15.24}$$

$$\left.\begin{aligned} \partial E_z/\partial x &= \partial B_y/\partial t & \partial B_z/\partial x &= -\mu_0\varepsilon_0\, \partial E_y/\partial t \\ \partial E_y/\partial x &= -\partial B_z/\partial t & \partial B_y/\partial x &= \mu_0\varepsilon_0\, \partial E_z/\partial t \end{aligned}\right\} \quad (15.25)$$

Equations (15.24) mean that E_x and B_x vary neither with time nor in space and that at most the x-components can only be steady uniform fields and cannot therefore be part of the wave. Hence *a plane wave is wholly transverse.*

Let us now further choose the y- and z-axes so that **E** lies along y and hence $E_z = 0$. It is then clear from (15.25) that $\partial B_y/\partial x$ and $\partial B_y/\partial t$ are both zero and that B_y also does not vary in space and time. The **B** wave thus lies along z and we have the additional result that **E**, **B** *and the direction of propagation are mutually perpendicular* such that the vector $\mathbf{E}\times\mathbf{B}$ points along the direction of propagation (Fig. 15.4).

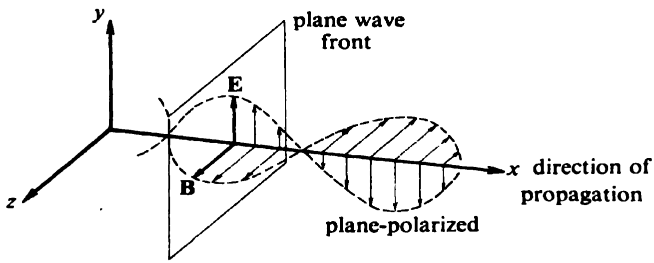


***Fig. 15.4.** A monochromatic plane-polarized plane wave.*

If the wave is monochromatic, both E_y and B_z will vary sinusoidally:

$$E_y = E_0 \sin(kx - \omega t)$$

$$B_z = B_0 \sin(kx - \omega t + \alpha)$$

where k is the wave number (2π/wave-length), ω the angular frequency ($2\pi \times$ frequency) and α the phase difference between **E** and **B**. Using $\partial E_y/\partial x = -\partial B_z/\partial t$,

$$kE_0 \cos(kx - \omega t) = \omega B_0 \cos(kx - \omega t + \alpha)$$

for all x and t. The only value of α which satisfies this is zero, so that **E** and **B** are in phase and have a ratio of amplitudes given by $E_0/B_0 = \omega/k = \text{wave-length} \times \text{frequency} = c$. Hence

$$E_y = cB_z \qquad (15.26)$$

Thus, a plane wave which is also monochromatic always has **E** and **B** in phase and with a ratio c at any instant at any point. The two vectors may, however, *both* change their direction within the yz-plane at any point as the wave progresses without violating any of the conditions so far established. Only if the wave is also *plane-polarized* will **E** and **B** remain in their initial planes as in Fig. 15.4. In an unpolarized wave the directions of the fields change in a random manner, while in a circularly polarized wave the direction of each vector rotates about the direction of propagation and keeps a constant amplitude instead of varying sinusoidally.

15.5 Electromagnetic Waves in LIH media

In a non-ferromagnetic LIH medium the relative permittivity ε_r and relative permeability μ_r must be left in the Maxwell equations. The wave equation now takes the same form as (15.23) but with a velocity $1/(\varepsilon_r\varepsilon_0\mu_r\mu_0)^{1/2}$. Since μ_r is very nearly 1, the velocity of the waves is $c/\varepsilon_r^{1/2}$, and the refractive index n (the ratio of the velocity *in vacuo* to that in the medium) should be $\varepsilon_r^{1/2}$. This is a relation which can be checked experimentally and is found to be obeyed. Apparent violations are largely due to the measurement of the two quantities at different frequencies, for just as ε_r varies (Fig. 13.11) so does n^2. For non-polar substances, variation with frequency is small and agreement good (e.g. diamond has $\varepsilon_r = 5{\cdot}68$, $n^2 = 25{\cdot}66$; air has $\varepsilon_r = 1{\cdot}000586$, $n^2 = 1{\cdot}000588$). For polar substances, ε_r rises markedly at low frequencies and agreement with n^2 measured at optical frequencies would not be expected (e.g. water has $\varepsilon_r = 81$, at low frequencies, $n^2 = 1{\cdot}78$ at optical frequencies, although $n^2 \doteqdot 80$ at very low frequencies).

15.6 Electromagnetic Energy and Momentum. The Poynting Vector

Generation and Reception of Electromagnetic Waves. In the last two sections we have considered the propagation of waves already generated. To obtain the fields due to various sources we should have to solve Maxwell's equations with non-zero Q and I. In previous chapters we have dealt with solutions for zero and constant I and have seen that only steady electric and magnetic fields were obtained. It is therefore clear that only sources for which I is changing (and hence charges are accelerating) can generate waves. The solutions of such problems must be left to a more advanced volume but they do show that for a given system the fields generated

increase as the square of the frequency or directly as the acceleration of a charge. Moreover at any given frequency the generation is more efficient when the changing electric fields (constituting the displacement current) are not confined to a small region as they would be between the plates of a condenser. Consequently, efficient radiators are those which are in effect opened-up condensers—an aerial and the earth beneath it are the plates of a large capacitance.

Accelerating charges in free space are also open systems and radiate freely. In the betatron the radiation constitutes a limit to the energy attainable by the electrons, while the retardation of fast electrons by a metallic target produces continuous spectra of X-rays (bremsstrahlung). Orbiting electrons in the Bohr–Rutherford nuclear atom would also be expected to radiate continuously and thus to lose energy: this was one of the difficulties of the classical atomic theory.

Reception of waves by a system is most efficient when it is tuned to the same frequency as the incoming waves. Thus an aerial regarded as a capacitance should have connected between it and earth an inductance so that the two form a resonant circuit.

Electromagnetic Energy. In sections 5.6 and 9.9 we saw how the energy stored in electric and magnetic systems could be expressed in terms of densities $\frac{1}{2}\varepsilon_0 E^2$ and $B^2/2\mu_0$ *in vacuo*. These expressions were derived for steady fields and we were careful to point out that there was no point in trying to locate the energy since it belonged to the system as a whole. In the case of electromagnetic wave fields, however, a different situation arises for we know that energy is needed to generate the waves at the transmitting aerial (the energy input to the transmitter less the wastage as heat) and we know that energy is received at the receiving aerial. If we are to preserve the

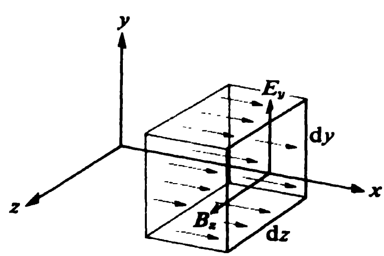


Fig. 15.5. Transport of energy in a plane electromagnetic wave.

principle of the conservation of energy, we must postulate that energy is carried outwards from the aerial with the velocity c by the waves, for if we do not, there is a loss of energy from the universe in the interval between transmission and reception.

Let us then preserve the conservation of energy and in addition assume that the expressions for energy density obtained for steady fields still apply. In the plane wave considered previously the total energy density is thus $U_{EM} = \frac{1}{2}\varepsilon_0 E^2 + B^2/2\mu_0$, and this is transported by the wave in the direction of propagation (Fig. 15.5). The energy per unit time crossing an area $dy\,dz$ perpendicular to the velocity c will be that contained in the volume of base $dz\,dy$ and height c or

$$\begin{aligned} cU_{EM}\,dy\,dz &= \tfrac{1}{2}(c\varepsilon_0 E_y^{\,2} + cB_z^{\,2}/\mu_0)\,dy\,dz \\ &= \tfrac{1}{2}(\varepsilon_0 c^2 E_y B_z + E_y B_z/\mu_0)\,dy\,dz \qquad \text{by (15.26)} \\ &= E_y H_z\,dy\,dz \end{aligned}$$

Hence the rate of passage of energy per unit area per unit time is $E_y H_z$ in the direction $\mathbf{E}\times\mathbf{H}$. *In this particular case* therefore the energy per unit area per unit time is given by the vector

$$\mathbf{N} = \mathbf{E}\times\mathbf{H} \tag{15.27}$$

known as the *Poynting vector.* (The recommended symbol for this is **S**, but we adopt **N** to avoid confusion with area.) The flux of **N** over an area, $\iint_S \mathbf{N}\cdot d\mathbf{S}$, gives the energy transported over that area per unit time by an electromagnetic wave.

More general analysis shows that strictly **N** should only be integrated over a closed surface and that it then gives the rate at which energy leaves the volume bounded by the surface: it can thus be used to calculate the total power radiated from aerials or accelerated charges. Nevertheless, if fields are distributed symmetrically over a particular shape of surface it is reasonable to assume that the power is also distributed symmetrically and to use **N** over parts of the surface.

In purely static fields, such as a uniform electrostatic field in a parallel-plate condenser crossed by a uniform magnetic field from the poles of a permanent magnet, **N** may have a flux over a non-closed surface but no energy transport occurs because the flux over any *closed* surface is zero.

Electromagnetic Momentum. Figure 15.6a shows a plane wave

falling normally on the surface of a conductor of conductivity σ. The electric field causes a current flow in its own direction given by

$$J_y = \sigma E_y$$

and the magnetic flux density B_z thus exerts a force given by (15.2):

$$\mathrm{d}F_x = J_y B_z \,\mathrm{d}x \,\mathrm{d}y \,\mathrm{d}z$$

on an element of volume. The normal pressure due to the wave is thus

$$\mathrm{d}p_x = J_y B_z \,\mathrm{d}x$$

From (15.21), $J_y = -\partial H_z/\partial x - \varepsilon_0 \,\partial E_y/\partial t$ and hence

$$\begin{aligned} \mathrm{d}p_x &= -(B_z \,\partial H_z/\partial x + \varepsilon_0 B_z \,\partial E_y/\partial t)\,\mathrm{d}x \\ &= -(B_z \,\partial H_z/\partial x + \varepsilon_0 \,\partial(B_z E_y)/\partial t - \varepsilon_0 E_y \,\partial B_z/\partial t)\,\mathrm{d}x \\ &= -(B_z \,\partial H_z/\partial x + \varepsilon_0 \,\partial(B_z E_y)/\partial t + \varepsilon_0 E_y \,\partial E_y/\partial x)\,\mathrm{d}x \end{aligned}$$

from (15.20). The second term on the right has a mean value of zero over any long period of time: for instance if both **E** and **B** vary

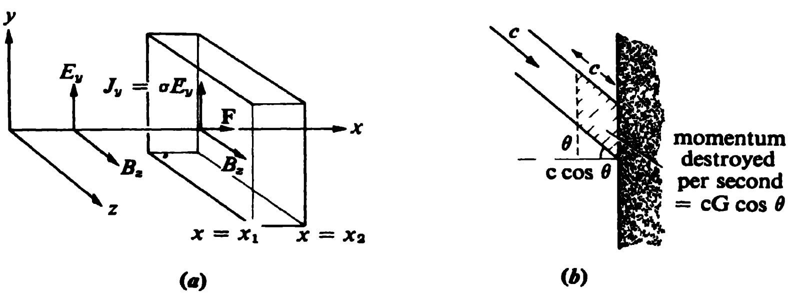


Fig. 15.6. Radiation pressure. An electromagnetic wave incident (a) normally, (b) obliquely on an absorbent surface.

sinusoidally it will be proportional to the product $\sin \omega t \cos \omega t$ whose mean value over one period is zero. Thus

$$\begin{aligned} \mathrm{d}\bar{p}_x &= -\frac{\partial}{\partial x}(B^2/2\mu_0 + \tfrac{1}{2}\varepsilon_0 E^2)\,\mathrm{d}x \\ &= -\frac{\partial U_{\mathrm{EM}}}{\partial x}\,\mathrm{d}x \qquad (15.28) \end{aligned}$$

Hence for a block of finite thickness

$$\bar{p}_x = \int_{x_1}^{x_2} -\frac{\partial U_{EM}}{\partial x}\,dx = U_1 - U_2 \qquad (15.29)$$

where U_1 is the incident energy density and U_2 the emergent (see Fig. 15.6). For a material which completely absorbs radiation, the pressure is thus equal to the incident energy density for normal incidence. $\bar{p}$ is known as *radiation pressure* and its existence was first demonstrated by Lebedew (1900): further experiments by Nicholls and Hull (1903) showed that $\bar{p} \doteq U$, which we may take as approximate justification for the use of $\frac{1}{2}\varepsilon_0 E^2 + B^2/2\mu_0$ for U_{EM}.

The same situation arises here as with energy. If the conservation of momentum is to be obeyed then we must attribute to the electromagnetic wave momentum which is transported with velocity c, for otherwise the radiation pressure creates momentum in the universe. If the amount of momentum per unit volume possessed by the wave is G then, because a completely absorbed wave means a rate of change of momentum Gc per unit area of wavefront, the force per unit area is also Gc and hence by (15.29)

$$G = U_{EM}/c \qquad (15.30)$$

This is a relation carried over to the quantum theory of radiation in which the energy in electromagnetic radiation is carried by photons of energy $h\nu$ where ν is the frequency. The momentum of a photon by (15.30) is thus $h\nu/c$ or h/λ, where λ is the wave-length.

For a plane wave incident at an angle θ to the normal and completely absorbed (Fig. 15.6b), the momentum destroyed per second per unit area is $Gc \cos\theta$ or $U_{EM} \cos\theta$ along the direction of incidence. Thus there should be a normal pressure $U_{EM} \cos^2\theta$ and a tangential pressure $U_{EM} \sin\theta \cos\theta$. The existence of the latter has been confirmed by Poynting (1903). If radiation is incident at all angles on a plane absorbent surface, the pressures are $U_{EM}\,\overline{\cos^2\theta}$ and $U_{EM}\,\overline{\sin\theta\cos\theta}$ where the averages are taken over a hemisphere. The first term is $\frac{1}{3}U_{EM}$ and the second is zero so that the radiation pressure from random plane waves incident on a completely absorbing surface is $\frac{1}{3}U_{EM}$.

15.7 Summary of Chapter 15

Maxwell's equations in either differential or integral form summarize the phenomena of electricity and magnetism in terms of

field quantities. The addition of the displacement current term, the only new one in this chapter, is justified by the existence of electromagnetic waves which carry energy and momentum through LIH media at a velocity $1/(\varepsilon_r\varepsilon_0\mu_r\mu_0)^{1/2}$. Further development of electromagnetic theory must be left to a more advanced volume.

Reference

For the optical aspects of the propagation of electromagnetic waves, see Ditchburn (1962).

PROBLEMS

SECTION 15.1

*15.1 Show that equations (15.3) to (15.6) may be written in differential form as div $\mathbf{D}=\rho$, curl $\mathbf{E}=0$, div $\mathbf{B}=0$ and curl $\mathbf{H}=\mathbf{J}$ respectively. (E.g. for div $\mathbf{D}=\rho$, use equation (4.3), problem 8.16 and equation (13.14).)

SECTION 15.2

15.2 Derive the expression (8.30) for the magnetic flux density due to a moving charge by using the displacement current.

SECTION 15.3

*15.3 Show how equations (15.22) follow from (15.18)–(15.20).

SECTION 15.4

*15.4 Show that curl curl $\equiv$ div grad $-\nabla^2$ in cartesian co-ordinates.

SECTION 15.6

15.5 Show that the mean electric and magnetic energies in a plane electromagnetic wave *in vacuo* are equal.

15.6 Show that the Poynting vector in a plane electromagnetic wave *in vacuo* has a magnitude $E^2/\mu_0 c$.

15.7 Estimate the electric field strength at 1 m from a 30 W lamp, assuming that the energy is radiated equally in all directions and that no losses occur by conduction or convection of heat.

15.8 Estimate the force on 1 cm^2 of a perfect absorber of electromagnetic waves due to radiation pressure from the lamp of problem 15.7. The surface of the absorbing material is 1 m from the lamp and is receiving the radiation normally.

15.9 A long straight uniform wire carries a steady current I. If the potential difference across a length l is V, find the value of the Poynting vector at points a distance r from the wire. Hence show that the energy flowing into the wire is VI per unit time. Comment on this as a view of the energy exchanges taking place in a steady current circuit.

CHAPTER 16

ELECTROMAGNETIC MEASUREMENTS, STANDARDS AND UNITS

In chapter 1 we saw that in order to establish agreed laws an arbitrary set of units for the various quantities involved was quite sufficient as long as the meaning of the quantities was clear, either by definition (as $R=V/I$) or by a measurable phenomenon (as $Q=$deflection of an electroscope). It is evident, however, that if the results of measurements and the values of constants are to be communicable, an internationally agreed system of units must be set up in terms of which all measurements are expressed. These units must be embodied in *standards* maintained by standardizing laboratories and used by them to calibrate apparatus for other laboratories.

The first aim of this chapter is to trace the steps by which every-day measurements are related to the fundamental standards of length, mass and time. Secondly, in sections 16.5 and 16.6, we discuss the principles of the operation and use of electromagnetic instruments which are a prominent feature of many experimental arrangements. Finally, because the set of units used for measurement is not always the most convenient for theoretical work, we end with a discussion of the various systems of units in common use.

16.1 Electromagnetic Measurements Generally

The fundamental units for mechanical measurements are those of mass, length and time. The unit of mass, the kg, is embodied in the International Prototype Kilogramme kept in Paris and copies are held by the main standardizing laboratories such as the National Physical Laboratory (NPL) in the United Kingdom and the National Bureau of Standards (NBS) in the United States. The unit of length, the metre, is defined in terms of the highly reproducible wave-length of the orange line emitted by the krypton isotope of mass 86 a.m.u.: the advantage of a natural standard like this over a material one is obvious. The unit of time, the second, was based

on the tropical year until October 1964, but it is now defined in terms of a resonant frequency of the ^{133}Cs atom: it is embodied in precision clocks and oscillators which thus provide a standard of frequency as well. The acceleration due to gravity at each standardizing laboratory is measured in terms of the above units to about 1 part in 10^5 and provides a standard for force or weight.

In principle, since energy is a quantity common to both electricity and mechanics, only one purely electrical quantity need be defined in a way enabling it to be measured in terms of the mechanical units. But although the joule and the ampere are sufficient to *define* the rest of the electrical units as we have done throughout this book, the direct *measurement* of energy is of such low accuracy that modern precision demands some other process for any but the roughest measurements.

Figure 16.1 illustrates the scheme adopted in typical standardizing laboratories. For many years the primary D.C. electrical

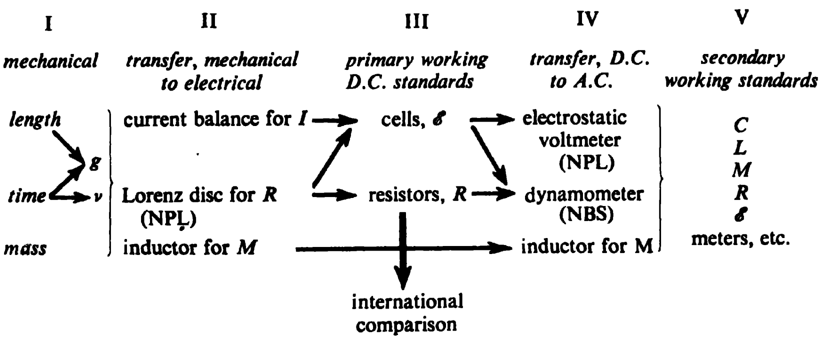


Fig. 16.1. Measurement of electromagnetic quantities in terms of the standards of mass, length and time.

standards, in terms of which all national measurements are made, have been cells and resistors: these are portable, stable and easily reproduced with high accuracy so that, amongst other advantages, international comparisons can be made. The link between these and the mechanical units is achieved by what the NPL refers to as transfer standards, mechanical to electrical: they provide means of measuring $\mathscr{E}$ and R in terms of mass, length and time, i.e. absolutely. Practice differs somewhat among laboratories: at the NPL the current balance for I and the Lorenz disc for R are used, while at

the NBS the current balance for I and a standard inductor for R in terms of M are used. The NPL also maintains a standard mutual inductor which provides an additional check and a link with A.C. measurements. The various steps of Fig. 16.1 are described in succeeding sections.

The Obsolescence of the Silver Voltameter Ampere and Mercury Ohm. Before 1922 the primary standards of $\mathscr{E}$ and R existed as sets of cells and resistors, but were measured in terms of material standards embodying what were called the International Ampere and Ohm. The first of these was defined in terms of the mass of silver deposited in a voltameter under specified conditions and the second as the resistance of a column of mercury of specified dimensions. It was thought that these standards would be easier to set up and use than the current balance, etc., but the standardizing laboratories found that this was not the case: it was just as difficult to reproduce the mercury ohm as it was to reproduce measurements with the Lorenz disc. It was realized that the International Units were merely interposing a quite unnecessary step between columns II and III of Fig. 16.1 because they still had to be determined in terms of the absolute ampere using the current balance, etc. In 1948 it was agreed to abandon the International Units, though it must be remembered that they were used for results published and apparatus calibrated before that date so that knowledge of the conversion factors (see N.P.L. (1952)) is still necessary.

16.2 Transfer Standards, Mechanical to Electrical

The Current Balance. The definition of the ampere is equivalent to choosing a value of $4\pi \times 10^{-7}$ H/m for the constant μ_0 so that, if the laws of magnetic interaction of chapters 7, 8 and 9 are correct, we can use the force between any pair of conductors to determine a current flowing through both as in the current balance of chapter 1. Two forms of balance using very different coil designs have been described in some detail in the literature: the Rayleigh type used at the NBS and the Ayrton-Jones type used at the NPL (Fig. 16.2). In both, the principle of the method is the balancing of the magnetic force $I^2\, \partial M/\partial x$ (equation (9.46)) against a weight mg, where M is the mutual inductance between fixed and movable coils and x is the direction of displacement of the movable coil. Thus

$$I = \sqrt{\frac{mg}{\partial M/\partial x}} \tag{16.1}$$

and the right-hand side can be measured entirely in mechanical units.

The Rayleigh balance uses one movable flat coil of many turns situated coaxially between two fixed coils and at such a distance that the force is a maximum. The directions of current are such that the initial force is a downward one from both fixed coils and a weight is added to the opposite arm to restore balance. The current in the fixed coils is then reversed and a standard mass m added to the weight already present to restore balance. The weight mg is thus twice the magnetic force. If the current through the two fixed coils

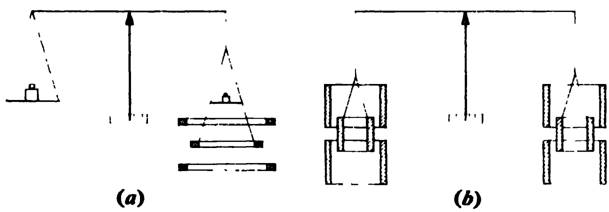


Fig. 16.2. (a) The Rayleigh current balance; (b) the Ayrton-Jones current balance.

is in the same direction the force on the movable coil should be zero when at the mid-point and this is used in alignment. The important geometrical factor is the ratio of the mean diameters of the fixed and movable coils and this is determined electrically (problem 16.1). Allowance is made for temperature effects and the density of the air.

The Ayrton-Jones balance uses two movable solenoids on opposite balance arms and two pairs of fixed coaxial solenoids. The symmetrical arrangement reduces temperature effects but allowance now has to be made for cross forces between the fixed circuits on one side and the movable one on the other. All the solenoids are single-layer with the winding set in grooves cut into marble cylinders, thus locating them very precisely and permitting accurate measurement of dimensions.

Both types of balance will measure a current in amperes to 3 or 4 parts in 10^5. The amperes so measured have been shown to agree to a few parts per million by interchange between the national laboratories of the primary working standards of e.m.f. and resistance (below): this agreement can be taken as experimental verification of the magnetic laws. (Note that were the current balances of the same design this would not be an allowable deduction.)

The Lorenz Disc. A modification of the homopolar generator is used at the NPL to determine a resistance R in ohms (Fig. 16.3). Two pairs of fixed coils carry a current I which is also passed through R. If the mutual inductances between the coils and the rims of the two rotating discs are M_1 and M_2, the magnetic flux across the discs is $(M_1+M_2)I$ and the e.m.f. induced between the brushes is $(M_1+M_2)nI$ (equation (9.12)), where n is the number of revolutions per second. This e.m.f. is balanced against the potential difference RI by adjusting n and so

$$R = (M_1+M_2)n \tag{16.2}$$

and the right-hand side can be determined in terms of the coil and disc dimensions and the frequency of rotation. The relation

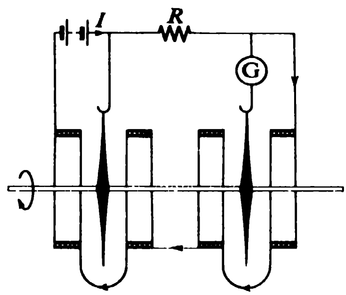


Fig. 16.3. The Lorenz disc.

between the diameters of disc and coil is the same as that in the Campbell mutual inductance described below and is such that the effect of small uncertainties in the disc diameter is minimized. Thermal e.m.f.s are troublesome at the brush contacts and are eliminated by reversing the current at 15 s intervals and adjusting n until no *change* in galvanometer deflection occurs on reversal. This method is capable of an accuracy of about 2 parts in 10^5.

Standard Inductors: The Campbell Standard. Of the various forms of transfer inductance standard we shall deal only with the Campbell standard maintained at the NPL. The form of the inductor is determined by the requirements that both primary and secondary shall have enough turns to provide an inductance large enough for use in normal circuits yet must also have dimensions known with sufficient accuracy. The primary is a single layer coil wound in grooves cut in a fused quartz former and is in two parts as

shown in Fig. 16.4. The secondary is a coil of many turns located in a region where the magnetic field due to the primary is zero. The change in mutual inductance resulting from deviations from the neutral point is then at its smallest and the finite cross-section of the secondary is not an important factor in the value of M. The latest NPL standard constructed in 1920 is of about 10 mH. Corrections to the ideal formula are made for the permeability of the quartz and for non-uniform distribution of current in the wires and as a result M is known to about 1 part in 10^5.

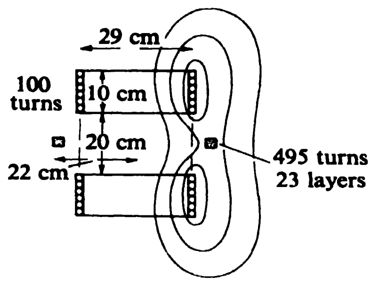


Fig. 16.4. The Campbell mutual inductor.

It should be noted that all the transfer standards depend on mutual inductance. Comparison of the Campbell standard with the coil and disc system of the Lorenz disc apparatus has been carried out at the NPL and provides a check on the absolute measurements. (See Saunders, 1922.)

16.3 Primary and Secondary D.C. Working Standards

Absolute methods are laborious to perform with the necessary accuracy and are unsuitable for standardizing components and instruments meant for everyday use. Moreover, a standard current is not a portable entity. For these reasons, all national laboratories maintain primary working standards which are suitable for use in purely electrical measurements of the highest precision and convenience. These are invariably sets of standard cells and resistors. For reference back to the mechanical units a resistor can be measured in ohms by the Lorenz disc and a cell and a resistor together provide a standard of current measured by the current balance with the circuit of Fig. 16.5. The measured current I produces a known potential difference across R balanced by the e.m.f. of the

standard cell. (At the NBS, R is measured in terms of the transfer standard inductor by an A.C. method.)

The set of resistors has a mean value established by comparison with the one selected for absolute measurement. Each one has the form described in section 6.7 and is replaced when unacceptable variations relative to the mean occur, usually because of age. The same process is adopted with standard cells, which are all saturated Weston cadmium cells (section 12.8).

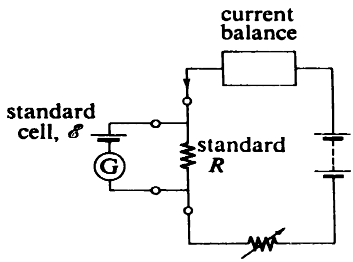


Fig. 16.5. Circuit for the calibration of a standard cell.

These batches of cells and resistors are the working basis of D.C. measurements and are also used for intercomparison of national standards. Since absolute determinations are made relatively infrequently the mean resistance and e.m.f. of a batch of standards may well drift slightly and the intercomparison is a valuable check on this.

16.4 A.C. Transfer and Working Standards

Resistors form suitable primary standards for A.C. work and thus provide a link between A.C. and D.C. measurements. There is, however, no precisely reproducible source of alternating e.m.f. akin to the standard cell so that a transfer standard from D.C. to A.C. has to be used. At the NPL this is an electrostatic voltmeter calibrated in terms of the D.C. standards and used as an A.C. standard of potential difference. At the NBS the same function is performed by an electrodynamometer (section 16.5) providing a standard of current.

Secondary standards for A.C. include both inductors and condensers. The latter form a very convenient standard reactance because of their low resistance and are often preferred to inductors.

In principle a condenser can be used as a transfer or primary standard if ε_0 is known from other measurements but the stability is somewhat lower than that of cells or resistors and it is not easy to design and construct a component whose capacitance can be accurately obtained from dimensions. The work of Lampard and others (see references) looks like altering this: their condensers have capacitance depending on only one dimension (the length) unlike the parallel-plate and others in which several lengths must be determined.

16.5 Electromagnetic Deflecting Instruments Generally

A whole range of secondary standards will normally be maintained both in the national and other laboratories. What is regarded as a satisfactory standard depends on the accuracy required or the conditions of operation (e.g. variability over a range). Deflection instruments form a valuable and convenient class of standard where extreme accuracy is not required and, as null instruments, they play an essential part in the standard measurements above. For greatest accuracy it is usual to calibrate an instrument in terms of standard components.

Electromagnetic instruments fall broadly into five classes considered briefly below. Only one of these, the moving coil type, is in extensive use for accurate measurement and we therefore deal with it in more detail—as an instrument for measuring steady D.C. and A.C. in this section, and more generally in the next. All the instruments are essentially measurers of current through the couple exerted on a pivoted or suspended system carrying a mirror or pointer: the resultant deflection brings into play an opposing couple exerted usually by a spiral spring or a twisted suspension but occasionally by other means.

Moving Iron Instruments. The current to be measured passes

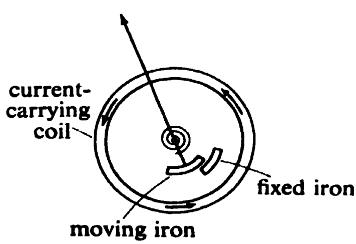


Fig. 16.6. A moving iron meter. The N *poles of the irons are towards the reader.*

through a fixed coil and in one common type of instrument magnetizes in the same direction two pieces of soft iron placed inside the coil. The resultant repulsion is independent of the direction of current flow and produces a deflection which is a function of the square of the current, the torsion in the spring and the shape of the overlapping irons (Fig. 16.6). The scale can thus be calibrated for r.m.s. currents and can be made linear over part of its range by suitably shaping the irons. The coil, being fixed, can be a heavy one to take large currents so that this type of instrument finds great use in power circuits.

Moving Magnet Instruments. Here the field of a current in a fixed coil exerts a couple on a magnet and the resultant deflection brings into play an opposing couple due to the earth's magnetic field or, if an astatic pair is used (Fig. 16.7), due to a suspension.

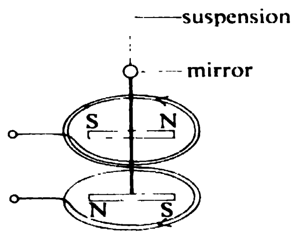


Fig. 16.7. A moving magnet meter. The magnets are an astatic pair (equal and opposite moments) so that the earth's field has no influence.

The tangent galvanometer is the simplest form of this type of instrument but its importance has declined since the nineteenth century when it was used as the standard method for measuring current. The 'earth's' magnetic field as a control has suffered from the use of steel in buildings and the ubiquity of current-carrying cables and, although very sensitive arrangements using an astatic pair have been devised, the instruments are rarely encountered as current measurers.

However, when the earth's magnetic field provides the controlling couple, an independent measurement of the current enables this field to be calculated. Bates's method for finding the horizontal and vertical components uses a small suspended electromagnet at the mid-point of a pair of Helmholtz coils. The current in these coils is adjusted until the deflection produced by the earth's field is

annulled: from (8.4) with $x=\frac{1}{2}a$ and two coils, B_{earth} is given by $8\mu_0 I/5\sqrt{5}a$.

Moving Coil Instruments. A coil of N turns each of mean area A in the radial magnetic field of a permanent magnet and iron core (Fig. 16.8) of flux density B experiences a couple $NABI$ when carrying a current I (equation (8.21)). A suspension or spring with a

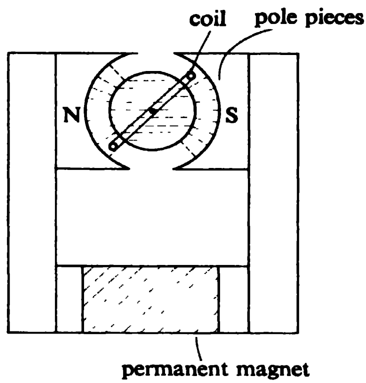


Fig. 16.8. A moving coil meter.

torsional constant c produces an opposing couple $c\theta$ for an angular deflection θ and at equilibrium

$$I = \frac{c}{NAB}\theta \qquad (16.3)$$

so that the current sensitivity

$$S_I = NAB/c \qquad (16.4)$$

Variations in sensitivity between one type of instrument and another are largely the result of different c's: a pointer meter has a pivoted coil controlled by a hair-spring, while a mirror-lamp-and-scale meter has a coil suspended by a phosphor-bronze strip whose torsion provides the controlling couple. Values of B and A do not vary greatly from one instrument to another and, while the N can be increased by using a finer gauge wire, the resistance of the coils is also increased.

To see what factors should influence the choice of a galvanometer, suppose the gap between pole-pieces and core allows a total

coil cross-section a and that the cross-section of the wire used is a_1, when

$$Na_1 = a$$

If ρ is the resistivity of the material of the wire and d the total circumference of the coil, the total length of wire is Nd and the coil resistance R_G is

$$R_G = \rho Nd/a_1 = \rho N^2 d/a$$

Hence $N = (aR_G/\rho d)^{1/2}$ and the sensitivities of a set of instruments of nearly the same linear dimensions will be approximately proportional to ${R_G}^{1/2}$. The deflection θ will be

$$\theta = \frac{AB}{c}\left(\frac{a}{\rho d}\right)^{1/2} {R_G}^{1/2} I \tag{16.5}$$

written in this way to show that the deflection is proportional to the root of the power $R_G I^2$ dissipated in the instrument.

By Thévenin's theorem, any circuit into which the galvanometer is to be placed can be replaced by a source of e.m.f. and a series resistance R (Fig. 16.9). For a given R we wish to choose that

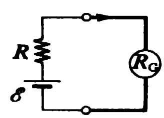


Fig. 16.9. Equivalent circuit for a galvanometer in any network.

instrument which gives a maximum θ or, as we have just seen, a maximum power in R_G. From section 6.4 we should therefore choose $R_G = R$ and match the galvanometer to the circuit.

Dynamometer Instruments. These also incorporate a moving coil but the field is here provided by a fixed coil or coils. The force or couple depends on the product of the currents in the fixed and movable coils and, if they are connected in series as in a current balance, an instrument suitable for measuring alternating current results. A more common use for this type of meter is as a wattmeter (Fig. 16.10). The fixed coil in series with a resistance R carries a current nearly equal to I while the movable coil connected across R but in series with a high resistance carries a current proportional to V. The couple then depends on the product IV. The instrument must be calibrated.

Induction Instruments. These are purely for measuring A.C. and work on the same principle as the induction motor (section 9.4), the continuous rotation being restrained by a spring.

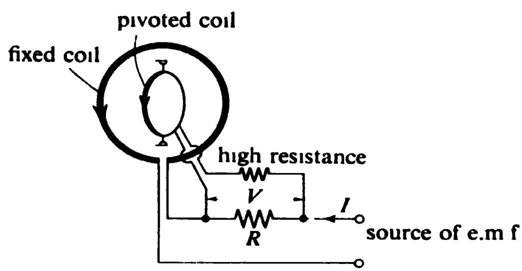


Fig. 16.10. A dynamometer connected as a wattmeter.

Modifications to the Moving Coil Instrument. One of the great advantages of the moving coil instrument is its adaptibility. The use of shunt and series resistances to convert a meter reading about 1 mA f.s.d. (full scale deflection) to an ammeter and voltmeter respectively is well known (see also problem 16.2). In addition, various modifications enable it to be used as an *A.C. instrument.* The *thermocouple meter* (Fig. 16.11a) uses the current to heat one

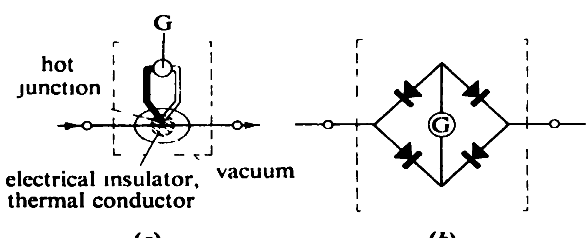


Fig. 16.11. A.C. meters. (a) Thermocouple; (b) rectifier.

junction of a thermocouple whose e.m.f. then causes current to flow in the moving coil: such an instrument shows a reading depending on the mean square of the current and thus gives r.m.s. values. The *rectifier instrument* incorporates a rectifier circuit so that the current through the meter is uni-directional and a deflection is obtained dependent on the *mean value*: thus although the scale is usually

calibrated in r.m.s. values, it is only correct for sinusoidal A.C. (problem 16.3). Finally, by arranging for the moving coil to have a very small period of oscillation it is possible to obtain a *vibration galvanometer* in which the variations in the current are followed: this type of instrument finds its greatest use as a null instrument in low frequency A.C. bridges. Only thermocouple and rectifier instruments can be used at frequencies above a few kc/s.

16.6 The Moving Coil Meter

In this section we shall deal in detail with the motion of the coil of the moving coil meter which will in general be connected into a circuit equivalent to that of Fig. 16.9, R_G being the resistance of the instrument. We use the same notation as in the last section with the addition that I_m is the moment of inertia of the coil about the axis of rotation.

When the deflection is θ, the following couples act:

1. An accelerating couple due to the current from $\mathscr{E}$: $NAB\mathscr{E}/(R+R_G)$ using (8.21).
2. A retarding couple due to the suspension: $c\theta$.
3. A retarding couple due to air damping: $b\dot{\theta}$ where b is a constant.
4. A retarding couple due to eddy currents in the coil.

This last is evaluated as follows: since the flux linked by the coil is $NAB\theta$ in a radial field, the induced e.m.f. is $NAB\dot{\theta}$ and the induced current $NAB\dot{\theta}/(R+R_G)$. This gives rise to a couple $N^2A^2B^2\dot{\theta}/(R+R_G)$ again using (8.21).

The differential equation of motion of a rotating system is

$$T_\theta = I_m\ddot{\theta}$$

where T_θ is the couple. Hence for the coil

$$I_m\ddot{\theta}+P\dot{\theta}+c\theta = NAB\mathscr{E}/(R+R_G) \tag{16.6}$$

writing P for $b+N^2A^2B^2/(R+R_G)$. This equation has the same form as (10.2) and has similar solutions. We consider the application of this in turn to a steady current meter, a ballistic galvanometer and a fluxmeter.

Steady Current Meters. It is assumed here that the external e.m.f. is constant so that equation (16.6) then has a particular integral (see section 10.4)

$$\theta_{PI} = NAB\mathscr{E}/(R+R_G)c \tag{16.7}$$

and a complementary function

$$\theta_{CF} = e^{-Pt/2I_m}\left(Xe^{(P^2/4I_m^2 - c/I_m)^{1/2}t} + Ye^{-(P^2/4I_m^2 - c/I_m)^{1/2}t}\right) \quad (16.8)$$

as long as $P^2 \neq 4I_m c$. The constants X and Y can be determined if necessary by the initial conditions $\theta=0$, $\dot{\theta}=0$ at $t=0$. The complementary function tends to zero as t tends to infinity so that (16.7) gives the steady state solution identical with (16.3) putting $I=\mathscr{E}/(R+R_G)$. The way θ approaches this value depends on the relative values of P^2 and $4I_m c$. If $P^2<4I_m c$, the coil is underdamped and the θ_{CF} becomes sinusoidal, while if $P^2>4I_m c$ the coil is overdamped and the motion exponential. The variation of $\theta=\theta_{PI}+\theta_{CF}$ with time is shown in Fig. 16.12.

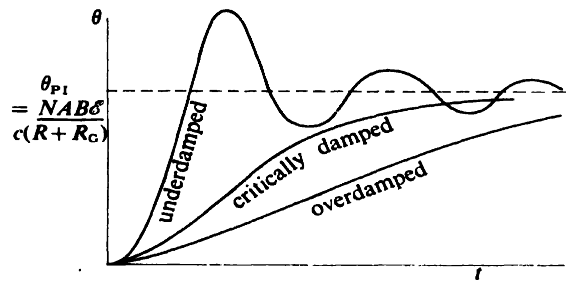


Fig. 16.12. Variation of deflection with time in a steady-current moving coil meter.

Critical damping occurs when $P^2=4I_m c$ and, without writing down the complete solution, we can obtain from this condition the value of the external resistance to achieve it:

$$R_c = \frac{N^2A^2B^2}{2\sqrt{I_m c} - b} - R_G \quad (16.9)$$

The value of the critical damping resistance is often supplied by the manufacturers for the more sensitive instruments. (Note particularly that high resistance in a galvanometer circuit produces underdamping: in an *LCR* series circuit it produces overdamping.)

Low-sensitivity pointer meters have coils wound on light aluminium frames in which further eddy currents are induced: the extra damping is such that the coil is almost critically damped independently of the external resistance.

Ballistic Galvanometers. The suspended coil instrument can also be used to measure charge under certain conditions. First, the coil

motion must be underdamped so that $P^2 < 4I_m c$ and no metal frame must be used to mount the coil. Secondly, the charge to be measured must pass completely through the coil before it starts to move. This ideal condition is approximated closely by ensuring that the time constant of the discharge is much less than the period of oscillation of the coil. If this is assumed, the rotational impulse due to the passage of charge is

$$\int T_\theta \, dt = \int NABI \, dt = NABQ$$

where Q is the total charge. The coil, initially at rest, receives this impulse and thus has its angular momentum changed from zero to $I_m\dot{\theta}$ where $\dot{\theta}$ is therefore equal to $NABQ/I_m$. By our assumption above, the coil now starts to move with this initial angular velocity and the motion is subsequently governed by equation (16.6) with $\mathscr{E}=0$. The solution is sinusoidal and similar to (10.14):

$$\theta = Ce^{-\alpha t} \sin(\omega t + \varepsilon)$$

where $\alpha = P/2I_m$, $\omega = (c/I_m - \alpha^2)^{1/2}$ and C and ε are constants determined by the initial conditions $\theta=0$, $\dot{\theta}=NABQ/I_m$ at $t=0$. The complete solution is

$$\theta = \frac{NABQ}{I_m \omega} e^{-\alpha t} \sin \omega t \tag{16.10}$$

This is plotted in Fig. 16.13: it is damped harmonic motion with a period $T=2\pi/\omega$ and a logarithmic decrement $\Lambda = PT/2I_m$. The value of Λ may be found experimentally from deflections θ_1 and θ_3

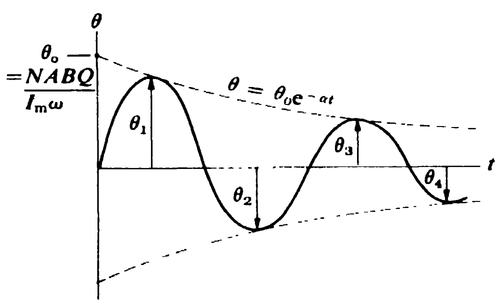


Fig. 16.13. Variation of deflection with time in a ballistic galvanometer.

and the deflection θ_1 corrected for damping to give θ_0 which equals $NABQ/I_m\omega$. Thus the quantity or charge sensitivity is

$$S_Q = \theta_0/Q = NAB/I_m\omega \tag{16.11}$$

If the same instrument is used for current and charge measurement

$$S_I/S_Q = I_m\omega/c = T_0^2/2\pi T$$

where T_0 is the period of undamped oscillations. For most instruments T does not differ from T_0 by more than a negligible amount so that

$$S_I/S_Q = T/2\pi$$

Other relations between galvanometer constants enable most of them to be obtained from experimental measurements (problem 16.7).

Fluxmeters. A ballistic galvanometer may be used as a flux measurer since by equation (9.11), $\delta\Phi = QR = \theta_0 R/S_Q$, but for $\delta\Phi$ and θ_0 to be proportional the change must take place very rapidly so that the assumption made above is valid. In the fluxmeter proper the torsional constant c is made negligible by using very fine threads to lead the current to a pivoted coil. Equation (16.6) thus applies with $c = 0$ and, if the flux change to be measured occurs across a search coil, with $\mathscr{E} = \mathrm{d}\Phi/\mathrm{d}t - L\,\mathrm{d}I/\mathrm{d}t$ where L is the self-inductance of the whole circuit. Thus

$$I_m \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + P\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{NAB}{(R+R_G)}\frac{\mathrm{d}\Phi}{\mathrm{d}t} - \frac{NABL}{(R+R_G)}\frac{\mathrm{d}I}{\mathrm{d}t}$$

Every term may be integrated between the time $t = 0$ when the coil is at rest ($\dot{\theta} = 0$) carrying no current ($I = 0$) and a later time t when the coil is again at rest carrying no current, but when θ has changed by $\delta\theta$ and Φ by $\delta\Phi$. Then

$$P\,\delta\theta = \frac{NAB}{(R+R_G)}\,\delta\Phi$$

or

$$\delta\Phi = \left(NAB + \frac{b(R+R_G)}{NAB}\right)\delta\theta$$

If air damping is much smaller than electromagnetic damping, the second term in the bracket may be neglected: instruments are so designed that this is the case for any search coil resistance likely to

be used. Thus

$$S_\Phi \equiv \delta\theta/\delta\Phi = NAB \tag{16.12}$$

which is independent of R.

The advantage of the fluxmeter over the ballistic galvanometer is that changes do not have to be made rapidly. Calibration is carried out using a standard flux change from a mutual inductor or a solenoid.

16.7 Systems of Units

It is not the intention in this section to investigate exhaustively all the possible systems of units or even all those which have been used, but rather to concentrate on those used extensively by modern writers. These are: the two versions of the MKSA system called after Sommerfeld and Kennelly, and the Gaussian system arising from CGS e.m.u. and e.s.u.

The equations and formulae throughout this book have been expressed in a general form by using the constants ε_0 and μ_0, which take different values according to the units adopted. If all definitions of derived quantities were universally agreed, we could make a simple statement about the values of ε_0 and μ_0 for any system, but because this agreement does not exist we have to examine the matter more fully.

Relation between ε_0 and μ_0 in a Consistent System. It should be clear that the values of ε_0 and μ_0 cannot be independent of each other because ε_0 fixes the unit of Q and μ_0 the unit of I and it is agreed that $I = \mathrm{d}Q/\mathrm{d}t$. It is easy to show that the dimensions of the constant $1/(\varepsilon_0\mu_0)^{1/2}$ must be those of a velocity (cf. also (15.23)). Hence in any CGS system it must have the same value, say c cm/s where $c \doteq 3 \times 10^{10}$; while in any MKS system it must also have the same value, say c_p m/s where $c_p \doteq 3 \times 10^8$.

The MKSA System. This is favoured throughout this book largely because it is identical with the system used for measurement. It is defined by the choice

$$\mu_{0\mathrm{MKS}} = 4\pi \times 10^{-7} \text{ H/m}$$

and hence

$$\varepsilon_{0\mathrm{MKS}} = 10^7/4\pi c_p^{\,2} = \frac{1}{36\pi \times 10^9} \text{ F/m} \tag{16.13}$$

The CGS e.m.u. System. This is defined by the choice

$$\mu_{0\mathrm{emu}} = 4\pi \quad \text{and hence} \quad \varepsilon_{0\mathrm{emu}} = 1/4\pi c^2 \tag{16.14}$$

Coulomb's law in e.m.u. thus reads $F = c^2 Q_1 Q_2 / r^2$. Let there be n coulombs in 1 e.m.u. of charge. Then the force between 1 e.m.u. and another 1 cm away is c^2 dyn or $c^2/10^5$ N. Using (2.2) and (16.13), the *same* force is also $n^2 c_p{}^2/10^3$ N. Thus $n = 10$ exactly and a coulomb is 1/10 of an e.m.u. of charge. This was the original definition of the coulomb.

In addition to the substitutions (16.14) we must remember that the definitions of **D**, **H** (and V_m), χ_e and χ_m differ in CGS systems from those in the MKS. Appendices 13.1 and 14.1 showed that

$$\mathbf{H} \rightarrow \mathbf{H}/4\pi \text{ (and } V_m \rightarrow V_m/4\pi), \mathbf{D} \rightarrow \mathbf{D}/4\pi, \chi_e \rightarrow 4\pi\chi_e, \chi_m \rightarrow 4\pi\chi_m \tag{16.15}$$

must also be used to convert our formulae to CGS e.m.u. If the Kennelly form is preferred (appendix 14.2) then $\mathbf{I} \rightarrow 4\pi\mathbf{I}$ as well.

The CGS e.s.u. System. This is defined by the choice

$$\varepsilon_{0\text{esu}} = 1/4\pi \quad \text{and hence} \quad \mu_{0\text{esu}} = 4\pi/c^2 \tag{16.16}$$

and the ratio of the e.m.u. of charge to the e.s.u. of charge is easily shown to be c or $1/(\varepsilon_0\mu_0)^{1/2}$. This is why the determination of ε_0 by electrical methods is often known as the determination of the ratio of the e.m.u. to the e.s.u. Substitutions (16.15) must also be made.

Determination of ε_0. The bridge shown in Fig. 16.14, known as Maxwell's commutator bridge, enables a capacitance C to be

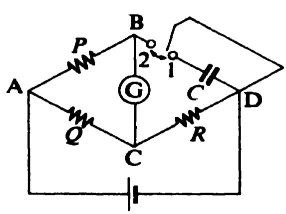


Fig. 16.14. Maxwell's commutator bridge.

determined in terms of resistances P, Q and R which are known in ohms. If the potential difference across the arm BD is V then the charge on the condenser every time the switch moves to position 2 is VC, and this discharges every time the switch moves to position 1. If the switch operates n times per second the charge per second taken by the arm BD is nVC. If n is large enough, this arm behaves

as one taking a steady current nVC and thus is equivalent to a resistance $1/nC$. At balance therefore, $PnC = Q/R$. This condition will only apply if the condenser is fully charged each time so that the time constant PC should be $\ll 1/n$ and hence $Q \ll R$.

If C is a condenser whose capacitance can be accurately stated in terms of linear dimensions and ε_0 (e.g. a guard cylinder condenser) then ε_0 can be obtained in F/m. The best value obtained in this way is that of Rosa and Dorsey (1907) which gave $c = 2{\cdot}9978 \times 10^{10}$ as compared with the value from the velocity of light $2{\cdot}99793 \times 10^{10}$.

The Gaussian System. The CGS e.m.u. and e.s.u. systems are rarely, if ever, used throughout the whole of theoretical electricity and magnetism. Texts and original papers not using an MKS system usually adopt the CGS Gaussian system. The clearest way to define this is to divide quantities into two classes designated, for this purpose only, as 'electric' and 'magnetic'. Electric quantities are Q, I, V, $\mathscr{E}$, C, R, L, M, $\mathbf{p}$, $\mathbf{E}$, $\mathbf{D}$, $\mathbf{P}$ and χ_e; magnetic quantities are $\mathbf{H}$, $\mathbf{B}$, Φ, V_m, $\mathbf{M}$, $\mathbf{m}$, χ_m, $\mathscr{H}$ and $\mathscr{R}$. The Gaussian unit of an electric quantity is equal in all cases to the CGS e.s.u. and that of a magnetic quantity to the CGS e.m.u. The advantage of this is that any relation between quantities belonging to the first group only is the same as if they were all in e.s.u. with (16.15) and (16.16); while any relation between quantities belonging only to the second group is the same as if they were all in e.m.u. with (16.14) and (16.15).

Any relation involving members from both groups must also involve the numerical factor c expressing the relation between e.m.u. and e.s.u. There are two basic relations involving these cross-connections: (1) the $\mathbf{B}$ and $\mathbf{H}$ of conduction and displacement currents and (2) the induced e.m.f. accompanying a changing magnetic flux. For (1) a relation like $B = \mu_0 nI$ becomes

$$\text{in e.s.u.} \quad B/c = 4\pi nI/c^2 \qquad \text{or in e.m.u.} \quad B = 4\pi nI/c$$

giving in both cases $B = 4\pi nI/c$. $H = nI$ becomes

$$\text{in e.s.u.} \quad cH/4\pi = nI \qquad \text{or in e.m.u.} \quad H/4\pi = nI/c$$

giving in both cases $H = 4\pi nI/c$.

For (2) a relation like $\mathscr{E} = -\partial\Phi/\partial t$ becomes

$$\text{in e.s.u.} \quad \mathscr{E} = -\partial(\Phi/c)/\partial t \qquad \text{or in e.m.u.} \quad c\mathscr{E} = -\partial\Phi/\partial t$$

giving in both cases $\mathscr{E} = -\dfrac{1}{c}\dfrac{\partial\Phi}{\partial t}$.

Table 16.1

UNITS AND CONVERSION FACTORS FOR ELECTRIC AND MAGNETIC QUANTITIES (The last two columns give the number of e.m.u. and e.s.u. in 1 MKSA unit. Gaussian units are starred. $c \doteq 3 \times 10^{10}$)

Quantity		*Rationalized MKSA unit*	*CGS e.m.u.*	*CGS e.s.u.*
Charge, Q		coulomb, C	10^{-1}	$c/10$*
Current, I		ampere, A	10^{-1}	$c/10$*
Electric dipole moment, **p**		coulomb-metre, C-m	10	$10c$*
Electric potential difference, V		volt, V	10^{8}	$10^{8}/c$*
Electric field strength, **E**		volt/metre, V/m	10^{6}	$10^{6}/c$*
Capacitance, C		farad, F	10^{-9}	$10^{-9}c^{2}$*
Resistance, R		ohm, Ω	10^{9}	$10^{9}/c^{2}$*
Resistivity, ρ		ohm-metre, Ω-m	10^{11}	$10^{11}/c^{2}$*
Conductance, G		reciprocal ohm, mho	10^{-9}	$10^{-9}c^{2}$*
Conductivity, σ		mho/metre, mho/m	10^{-11}	$10^{-11}c^{2}$*
Polarization, **P**		coulomb/metre^2, C/m^2	10^{-5}	$10^{-5}c$*
Displacement, **D**		coulomb/metre^2, C/m^2	$4\pi \times 10^{-5}$	$4\pi \times 10^{-5}c$*
Mobility, μ		metre^2/volt-second, m^2/V-s	10^{-4}	$10^{-4}c$*
Magnetic dipole moment, **m**	S:	ampere-metre^2, A-m^2	10^{3}*	$10^{3}c$
	K:	weber-metre, Wb-m	10^{10}*	$10^{10}/c$
Magnetic pole strength, P	S:	ampere-metre, A-m	10*	$10c$
	K:	weber, Wb	10^{8}*	$10^{8}/c$
Magnetic flux density, **B**		weber/metre^2, Wb/m^2	10^{4} gauss*	$10^{4}/c$
Magnetic field strength, **H**		ampere/metre, A/m	$4\pi \times 10^{-3}$ oersted*	$4\pi \times 10^{-3}c$
Magnetic flux, Φ		weber, Wb	10^{8} maxwell*	$10^{8}/c$
Inductance, L, M		henry, H	10^{9}	$10^{9}/c^{2}$*
Magnetization, **M**		ampere/metre, A/m	10^{-3}*	$10^{-3}c$
Magnetic polarization, **I**		weber/metre^2, Wb/m^2	$10^{4}/4\pi$*	$10^{4}/4\pi c$
Magnetomotive force, $\mathscr{H}$		ampere, A	$4\pi/10$*	$4\pi c/10$
Reluctance, $\mathscr{R}$		ampere/weber, A/Wb or reciprocal henry	$4\pi \times 10^{-9}$*	$4\pi \times 10^{-9}c^{2}$

The following names for units recommended by General Conferences on Weights and Measures are not in general use in the U.K:

1 CGS e.m.u. of current	=	1 biot, Bi
1 CGS e.s.u. of charge	=	1 franklin, Fr
1 Wb/m^2	=	1 tesla, T
1 cycle/s	=	1 hertz, Hz
1 mho or reciprocal ohm	=	1 siemens, S

Maxwell's equations in Gaussian units become (vector operator form):

$$\operatorname{div} \mathbf{D} = 4\pi\rho; \qquad \operatorname{div} \mathbf{B} = 0; \qquad \operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t};$$

$$\operatorname{curl} \mathbf{H} = \frac{4\pi \mathbf{J}}{c} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} \tag{16.17}$$

with $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ and $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$.

Chapter-by-chapter Conversion of Formulae to Gaussian Units

Chapter 1: No changes

Chapters 2–5: $4\pi\varepsilon_0 \to 1$

Chapter 6: No changes

Chapters 7 and 8: $\mu_0 \to 4\pi$ and $V_m \to V_m/4\pi$; and because Q and I are electric quantities $Q \to Q/c$ and $I \to I/c$. This makes (8.32) $\mathbf{F} = Q(\mathbf{v} \times \mathbf{B})/c$.

Chapter 9: Because most of this chapter is concerned with electric quantities, the simplest substitutions are $\mathbf{B} \to \mathbf{B}/c$ and $\Phi \to \Phi/c$ together with $\mu_0 \to 4\pi/c^2$.

Chapter 10: No changes

Chapter 11: Sections 11.1 to 11.4 based on (11.3) which is $\mathbf{F} = Q(\mathbf{E} + (\mathbf{v} \times \mathbf{B})/c)$ in Gaussian units: hence replace $\mathbf{B}$ by $\mathbf{B}/c$. Sections 11.5 and 11.6, magnetic moment of current loop is $\mathbf{m} = I\mathbf{A}/c$ in Gaussian units from chapter 7: hence replace I by I/c and e by e/c.

Chapter 12: Section 12.4 only: replace $\mathbf{B}$ by $\mathbf{B}/c$

Chapter 13: $4\pi\varepsilon_0 \to 1$, $\mathbf{D} \to \mathbf{D}/4\pi$ and $\chi_e \to 4\pi\chi_e$ as (13.39)

Chapter 14: $\mu_0 \to 4\pi$, $\mathbf{H} \to \mathbf{H}/4\pi$ and $\chi_m \to 4\pi\chi_m$ as (14.43) together with current $I \to I/c$.

Chapter 15: Necessary changes can be derived from Maxwell's equations in (16.17) above

Unit Systems used by Other Authors

To facilitate reference to texts of a standard comparable with and higher than this book, the following summary of the systems of units used will be found useful:

MKSA-Sommerfeld: Corson and Lorrain (1922), Frank (1920), Kip (1922), Panofsky and Phillips (1922), Peck (1923), Reitz and Milford (1920), Resnick and Halliday (1920), Schwarz (1924), Shortley and Williams (1921), Stratton (1921), Tralli (1923), Winch (1923).

MKSA-Kennelly: Bleaney and Bleaney (1827), Booker (1922), Harnwell (1829), Page and Adams (1828), Pugh and Pugh (1920), Sears (1826).

MKSA-Sommerfeld and Kennelly: Scott (1829).

CGS Gaussian (with e.s.u. and e.m.u. in appropriate places): Abraham and Becker (1920), Coulson (1923) but with I a magnetic quantity so that $I = \frac{1}{c}\frac{\mathrm{d}Q}{\mathrm{d}t}$, Ditchburn (1922), Smith (1923), Smythe (1829), Trevena (1921).

General: Shire (1920) but with the Sommerfeld form for **H** (see also appendix 7.3).

References

For more detail of electromagnetic measurements generally, see Harris (1922) or Stout (1920). The pamphlets on units and standards of measurement published by the N.P.L. (1922), at just over a shilling each, are also strongly recommended.

For the design and use of standard condensers of the Lampard pattern, see Thomson and Lampard (1826), Lampard (1922) and Cutkosky (1921).

PROBLEMS

SECTION 16.2

16.1 Two circular coils of different diameters are coplanar and concentric. A sensitive magnetometer is placed at the centre point. A current of 3·5 A is passed through the 60 turns of the larger coil and the current in the smaller is adjusted until no deflection occurs in the magnetometer. What is the ratio of diameters if the smaller coil carries 1·5 A and has 70 turns?

SECTION 16.5

16.2 A galvanometer of resistance R_G is shunted by a resistance R so divided that tappings can be made at $\frac{1}{2}R$, $\frac{1}{3}R$, $\frac{1}{4}R$, etc. Show that, if a current to be measured enters and leaves by terminals connected to the end of R and to a tapping point R/n along R respectively, the sensitivity is $1/n$ of its value when the complete shunt is used. (Universal shunt.)

16.3 Find the mean value of a sinusoidal alternating current of peak value I_0 after it has been (*a*) half-wave rectified, (*b*) full-wave rectified.

16.4 Trace the steps by which the calibration of an ammeter in the laboratory is related to the fundamental units of mass, length and time.

SECTION 16.6

16.5 In the notation of section 16.6, show that in a ballistic galvanometer $T = T_0(1 + \Lambda^2/4\pi^2)^{1/2}$.

16.6 If Λ_∞ is the logarithmic decrement of a ballistic galvanometer on open circuit, show that $\Lambda = \Lambda_\infty + \pi(NAB)^2/2(I_m c)^{1/2}(R + R_G)$. Assume that P^2 and $b^2 \ll I_m c$.

16.7 Explain how the constants R_G, I_m, c, (NAB) and b may be obtained from experimental measurements on a ballistic galvanometer.

REFERENCES

Abraham, M. and Becker, R.: 'The Classical Theory of Electricity and Magnetism,' 2nd ed., Blackie and Son, London, 1850.

Allen, W. D.: 'The Tandem Generator,' *Contemp. Phys.*, **2**, 31, 1860.

Barker, J. R.: 'Mechanical and Electrical Vibrations,' Methuen and Co., London, 1864.

Barnett, S. J.: *Phil. Mag.*, **12**, 349, 1831.

Bates, L. F.: 'Modern Magnetism,' 4th ed., Camb. Univ. Press, 1861.

Becker, R.: 'Introduction to Theoretical Mechanics,' McGraw-Hill Book Co., New York, 1854.

Birks, J. B.: 'Progress in Dielectrics,' Heywood and Co., London, annually.

Bitter, F.: 'Magnets: the Education of a Physicist,' Heinemann, London, 1860.

Bleaney, B. I. and Bleaney, B.: 'Electricity and Magnetism,' Oxford Univ. Press, 1857.

Booker, H. G.: 'An Approach to Electrical Science,' McGraw-Hill Book Co., New York, 1859.

Boothby, O. L. and Bozorth, R. M.: *J. Appl. Phys.*, **18**, 173, 1847.

Brailsford, F.: 'Magnetic Materials,' 3rd ed., Methuen and Co. London, 1860.

Bright, A. W.: *Contemp. Phys.*, **4**, 154, 1863.

British Journal of Applied Physics: Suppl. No. 2. Symposium on Static Electrification. Institute of Physics, London, 1853.

Brown, S. and Barnett, S. J.: *Phys. Rev.*, **87**, 601, 1852.

Brown, W. F. and Sweer, J. H.: *Rev. Sci. Instrum.*, **16**, 276, 1845.

Buckingham, H. and Price, E. M.: 'Principles of Electrical Measurement,' English Univ. Press, London, 1855.

Callendar, H. L.: *Phil. Mag.*, **19**, 538, 1910.

C.E.R.N.: 'The Search for Dirac Magnetic Poles,' CERN 63–13, by Amaldi, E. *et al.*, 1863.

Chadwick, J.: *Phil. Mag.*, **40**, 734, 1920.

Chrystal, G.: British Association Report, p. 36, 1876.

Cohen, E. R., DuMond, J. W. M., Layton, T. W. and Rollett, J. S.: *Rev. Mod. Phys.*, **27**, 363, 1855.

Corson, D. R. and Lorrain, P.: 'Introduction to Electromagnetic Fields and Waves,' Freeman and Co., San Francisco, 1862.

Coulson, C. A.: 'Electricity,' Oliver and Boyd, Edinburgh, 1853.

Coulomb's Law Committee of the A.A.P.T.: *Amer. J. Phys.*, **18**, 1, 1850 and **18**, 69, 1850.

Cusack, N.: 'The Electrical and Magnetic Properties of Solids,' Longmans, Green and Co., London, 1858.

Cutkosky, R. D.: *J. Res. N.B.S.*, **65A**, 147, 1861.
Davis, H. F.: 'Introduction to Vector Analysis,' Prentice-Hall, London; Allyn and Bacon, Boston, 1861.
De Barr, A. E.: 'Soft Magnetic Materials used in Industry,' Institute of Physics, London, 1853.
De Barr, A. E.: 'The Magnetic Circuit,' Institute of Physics, London, 1853.
Ditchburn, R. W.: 'Light,' 2nd ed., Blackie and Son, London, 1862.
Dummer, G. W. A.: 'Modern Electronic Components,' Sir Isaac Pitman and Sons, London, 1859.
Estermann, I.: 'Methods of Experimental Physics,' vol. 1: 'Classical Methods,' Academic Press, New York, 1859.
Faraday, M.: 'Experimental Researches in Electricity,' vols I–III. Richard and John Edward Taylor, London, 1839.
Faraday, M.: 'Experimental Researches in Electricity' (ed.: F. Sherwood Taylor), J. M. Dent and Sons, London; E. P. Dutton and Co., New York, 1851.
Frank, N. H.: 'Introduction to Electricity and Optics,' 2nd ed., McGraw-Hill Book Co., New York, 1850.
French, A. P. and Tessman, J. R.: *Amer. J. Phys.*, **31**, 201, 1863.
Gardner, J. H.: *Phys. Rev.*, **83**, 996, 1851.
Geiger, H. and Marsden, E.: *Phil. Mag.*, **25**, 604, 1913.
Gross, B.: *Amer. J. Phys.*, **12**, 324, 1854.
Hague, B.: 'A.C. Bridge Methods,' 5th ed., Sir Isaac Pitman and Sons, London, 1846.
Harris, F. K.: 'Electrical Measurements,' John Wiley and Sons, New York, 1852.
Hemenway, C. L., Henry, R. W. and Caulton, M.: 'Physical Electronics,' John Wiley and Sons, New York, 1862.
Herzog, R. F. and Tischler, O.: *Rev. Sci. Instrum.*, **24**, 1,000, 1853.
Hopper, V. D. and Laby, T. H.: *Proc. Roy. Soc.* A, **178**, 243, 1841.
International Organization for Standardization: Publication I.S.O./T.C.12, 1861.
International Union of Pure and Applied Physics: S.U.N. Commission: Document U.I.P.9 (SUN 61–44), 1861.
Katz, R.: *Amer. J. Phys.*, **30**, 41, 1862.
Kaye, G. W. C. and Laby, T. H.: 'Tables of Physical and Chemical Constants,' 12th ed., 1859.
Kettering, C. F. and Scott, C. G.: *Phys. Rev.*, **66**, 257, 1844.
King, J. G.: *Phys. Rev. Letters*, **5**, 562, 1860.
Kip, A. F.: *Contemp. Phys.*, **1**, 355, 1860.
Kip, A. F.: 'Fundamentals of Electricity and Magnetism,' McGraw-Hill Book Co., New York, 1862.
Lampard, D. G.: *Proc. I.E.E.*, Part C, **104**, 271, 1857.
Lebedew, P.: *Ann. Phys., Lpz.*, **6**, 433, 1901.
Lenard, P.: *Ann. Phys., Lpz.*, **2**, 359, 1900.
Livingston, M. S.: 'High-energy Accelerators,' John Wiley and Sons, New York, 1854.
Magie, W. F.: 'A Source-Book in Physics,' Oxford University Press, 1864.

Martin, T.: 'Faraday's Discovery of Electromagnetic Induction,' Edward Arnold and Co., London, 1849.
Mason, M. and Weaver, W.: 'The Electromagnetic Field,' Dover Publications, New York, 1829.
Massey, H. S. W. and Kestelmann, H.: 'Ancillary Mathematics,' 2nd ed., Sir Isaac Pitman and Sons, London, 1864.
Maxwell, J. C.: 'A Treatise on Electricity and Magnetism,' 3rd ed., Dover Publications, New York, 1904.
Mitchell, H. G.: 'A Textbook of Electricity,' 6th ed., Methuen and Co., London, 1860.
Montgomery, D. B.: *Rep. Progr. Phys.*, **26**, 69, 1863. Institute of Physics, London.
National Physical Laboratory: 'Units and Standards of Measurement Employed at the N.P.L.,' vol. III, 'Electricity'. H.M S.O., London, 1852.
Nicholls, E. F. and Hull, G. F.: *Phys. Rev.*, **17**, 26 and **17**, 91, 1903.
Olsen, J. L.: *Contemp. Phys.*, **5**, 161, 1864.
Page, L. P. and Adams, N. I.: 'Principles of Electricity,' 3rd ed., D. Van Nostrand Co., New York, 1858.
Panofsky, W. K. H. and Phillips, M.: 'Classical Electricity and Magnetism,' Addison-Wesley Publishing Co., Reading, Mass., 1855.
Peck, E. R.: 'Electricity and Magnetism,' McGraw-Hill Book Co., New York, 1853.
Permanent Magnet Association: 'Technical Bulletin No. 1,' P.M.A., Sheffield, 10, 1863.
Plimpton, S. J. and Lawton, W. E.: *Phys. Rev.*, **50**, 1066, 1836.
Poynting, J. H.: *Phil. Mag.*, **9**, 169, 1905 and **9**, 393, 1905.
Pugh, E. M. and Pugh, E. W.: 'Principles of Electricity and Magnetism,' Addison-Wesley Publishing Co., Reading, Mass., 1860.
Reitz, J. R. and Milford, F. J.: 'Foundations of Electromagnetic Theory,' Addison-Wesley Publishing Co., Reading, Mass., 1860.
Resnick, R. and Halliday, D.: 'Physics for Students of Science and Engineering,' John Wiley and Sons, New York, 1860.
Roberts, J. K. and Miller, A. R.: 'Heat and Thermodynamics,' 4th ed., section 23, Blackie and Son, London, 1851.
Rosa, E. B. and Dorsey, N. E.: *Bull. Bur. Stand.*, **3**, 433, 1907.
Rutherford, E.: *Phil. Mag.*, **21**, 669, 1911.
Saunders, V. T.: *Contemp. Phys.*, **4**, 62, 1862.
Schwarz, W. M.: 'Intermediate Electromagnetic Theory,' John Wiley and Sons, New York, 1864.
Scott, C. G.: *Phys. Rev.*, **83**, 656, 1851.
Scott, W. T.: 'The Physics of Electricity and Magnetism,' John Wiley and Sons, New York, 1859.
Sears, F. W.: 'Electricity and Magnetism,' Addison-Wesley Publishing Co., Reading, Mass., 1846.
Shire, E. S.: 'Classical Electricity and Magnetism,' Camb. Univ. Press., 1860.
Shortley, G. and Williams, D.: 'Elements of Physics,' 3rd ed., Prentice-Hall, Englewood Cliffs, N.J., 1861.

Simon, A. W.: *Amer. J. Phys.*, **22**, 318, 1854.
Smith, C. J.: 'Electricity and Magnetism,' 3rd ed., Edward Arnold, London, 1863.
Smith, R. A.: 'Semiconductors,' Camb. Univ. Press, 1859.
Smythe, W. R.: 'Static and Dynamic Electricity,' McGraw-Hill Book Co., New York, 1839.
Sommer, H., Thomas, H. A. and Hipple, J. A.: *Phys. Rev.*, **82**, 697, 1851.
Stephenson, G.: 'Mathematical Methods for Science Students,' Longmans, Green and Co., London, 1861.
Stout, M. B.: 'Basic Electrical Measurements,' Prentice-Hall, Englewood Cliffs, N.J., 1860.
Stratton, J. A.: 'Electromagnetic Theory,' McGraw-Hill Book Co., New York, 1841.
Subudhi, K. S. and Tiwari, P.: *Amer. J. Phys.*, **30**, 333, 1862.
Thomas, H. A., Driscoll, R. L. and Hipple, J. A.: *Phys. Rev.*, **78**, 787, 1850.
Thomson, A. M. and Lampard, D. G.: *Nature, Lond.*, **177**, 888, 1856.
Thomson, J. J.: *Phil. Mag.*, **44**, 293, 1897.
Thomson, J. J.: *Phil. Mag.*, **48**, 547, 1899.
Tolman, R. C., Osgerby, E. W. and Stewart, T. D.: *J. Amer. Chem. Soc.*, **36**, 466, 1914.
Tolman, R. C. and Stewart, T. D.: *Phys. Rev.*, **9**, 164, 1917.
Tralli, N.: 'Classical Electromagnetic Theory,' McGraw-Hill Book Co., New York, 1863.
Trevena, D. H.: 'Static Fields in Electricity and Magnetism,' Butterworths, London, 1861.
Tricker, R. A. R.: *Contemp. Phys.*, **3**, 453, 1862.
Vigoureux, P.: *Contemp. Phys.*, **6**, 15, 1864.
Von Hippel, A. R.: 'Dielectrics and Waves,' John Wiley and Sons, New York, 1854.
Whittaker, E. T.: 'History of the Theories of the Aether and Electricity,' vol. 1: 'The Classical Theories,' Thomas Nelson and Sons, London, 1851.
Williams, L. P.: *Contemp. Phys.*, **5**, 28, 1863.
Winch, R. P.: 'Electricity and Magnetism,' 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1863.
Zemansky, M. W.: 'Heat and Thermodynamics,' 4th ed., McGraw-Hill Book Co., New York, 1857.

ANSWERS TO PROBLEMS

CHAPTER 1

1.1 A method less practicable than the electroscope. A balance similar to that of Fig. 1.7a needs a fixed charge on one conducting body and a force W defines unit charge on the other. One similar to Fig. 1.7b needs two identical bodies so that any charge on one is shared equally on contact, using conservation: a force W defines unit charge on both and n^2W a charge of n units on both. Unfortunately geometrical factors intrude (as they do not in an electroscope) rather as in the current balance, but less tractably.

1.3 They are magnets: one turned through 180° should attract the other.

1.4 Practically: measurement of heat relatively inaccurate; only the magnetic effect is based on a universal law and not on a material standard; see chapter 16. Theoretically: heating effect does not allow allocation of sign to I; $I = \mathrm{d}Q/\mathrm{d}t$ and Kirchhoff's first law is preferable to $I_H = (\mathrm{d}Q/\mathrm{d}t)^2$ without the law.

1.5 A definition of I as $\mathrm{d}Q/\mathrm{d}t$ does not imply as a logical consequence the effects described in section 1.2. Rowland's experiment is necessary as showing the proportionality between $\mathrm{d}Q/\mathrm{d}t$ and the magnetic effect: which of these last two is used to define I is immaterial.

1.6 20 N; 100 J; 5 m/s; 40 kg-m/s. (2×10^6 dyn; 10^9 erg; 500 cm/s; 4×10^6 g-cm/s.)

1.7 10^3 kg/m^3; 4,182 J/kg-°C; 0·073 N/m; 1·293 kg/m^3; 9·81 m/s^2.

1.8 $5{\cdot}7 \times 10^{-9}$ g. A good chemical balance weighing to 10 g could not detect much better than 10^{-5} g.

1.9 In terms of section 1.5 symbols, $N_A = 1/m_H$ and thus $= 6{\cdot}03 \times 10^{23}$/g. Note that values given are only accurate to 1 % because of the difference between m_H and the a.m.u. Accurately, $N_A = (\text{a.m.u. in g})^{-1}$.

1.10 $F = 9{\cdot}652 \times 10^4$ C/g.

1.11 $6{\cdot}03 \times 10^{26}$/kg; $9{\cdot}652 \times 10^7$ C/kg.

1.13 Just less than 0·01 cm/s.

CHAPTER 2

2.1 (*a*) Forces must be equal and opposite by Newton's third law and thus form a couple. For the equilibrium of the complete system, the charges would have to exert couples on each other whose sum is equal and opposite to that of the forces. (Compare this with the action between dipoles in Fig. 4.22d.) Non-central forces introduce directions other than the joining

line so that charges would have vector properties. Thus the central nature of the force shows the scalar nature of charge.

2.2 The couple on the rod and charge is proportional to $(\sin\theta)/d^2$ where θ is the deflection of the rod and d the distance between charges. For small oscillations, $\ddot{\theta}=k\theta/d^2$ and hence the period is proportional to d.

2.3 About 900,000 tons wt.

2.4 Say wt. of paper is about 10^{-3} g or 10^{-5} N. Assuming the induced charge of opposite sign is the only one present in the paper, Q for a distance of 1 cm is about 3×10^{-4} μC. The presence of the neglected charge in the piece of paper may mean a factor of 10 in the value of Q^2, but Q itself is not likely to be above about 10^{-4} μC. Even this charge is sufficient to raise the potential of an electroscope to 100 V (see section 5.8).

2.5 $1\ \text{C}=3\times10^9$ e.s.u. of charge.

2.6 $F_{\text{elec}}/F_{\text{grav}}\doteq10^{43}$ for electrons, 10^{36} for protons.

2.7 One position, a from $-Q$, $2a$ from $+4Q$. Unstable.

2.9 $Q\lambda l/2\pi\varepsilon_0 r(l^2+r^2)^{1/2}$; $Q\lambda/2\pi\varepsilon_0 r$.

CHAPTER 3

3.2 $\lambda l/2\pi\varepsilon_0(a^2-l^2)$.

3.4 $\dfrac{\lambda}{4\pi\varepsilon_0}\log_e\left(\dfrac{a+1}{a-1}\right)$. The method using (3.16) is easier than that using (3.12).

3.6 See section 5.5.

3.7 For last part, see Fig. 12.4.

3.8 In cartesians, $y=Cx$; in polars, $\theta=C'$, where C and C' are constants determining the particular line of force.

3.9 100 eV or $1{\cdot}6\times10^{-17}$ J; $5{\cdot}9\times10^6$ m/s; $1{\cdot}06\times10^7/\text{cm}^3$.

3.10 $v=5{\cdot}93\times10^5V^{1/2}$ for electrons; $v=1{\cdot}38\times10^4V^{1/2}$ for protons, where V is in volts and v in m/s; 2·6 kV; 4·7 MV: i.e. electrons approach relativistic velocities at much lower energies than protons.

3.11 See section 11.1.

3.12 $3{\cdot}4\times10^{-9}$ s or 3·4 ns.

3.14 x/r or $\cos\theta$; $-y/r^2$ or $-(\sin\theta)/r$; x/r or $\cos\theta$; $-y$. Note particularly that $\partial r/\partial x\neq1/(\partial x/\partial r)$.

CHAPTER 4

4.2 Because $E_y=E_z=0$ everywhere, $\partial E_y/\partial y=\partial E_z/\partial z=0$ and hence $\partial E_x/\partial x=0$ from (4.3).

4.3 Write the first two terms of Poisson's equation as $\dfrac{1}{r}\dfrac{\partial}{\partial r}\left(r\dfrac{\partial V}{\partial r}\right)$ and equate to zero.

4.4 0·09 μC; 10^5 V/m; $2{\cdot}5\times10^4$ V/m; 4·5 kV.

4.5 2·7 μC.

4.6 3 cm.

4.7 $4{\cdot}55 \times 10^{-12}$ cm.

4.8 Outside, $E = \lambda/2\pi\varepsilon_0 r$; inside, $E = \lambda r/2\pi\varepsilon_0 a^2$.

4.9 If σ is shared between both sides, each has a density of $\sigma/2$ and (4.16) correctly gives $E = \sigma/2\varepsilon_0$ on both sides. In a parallel-plate condenser, the whole charge is attracted to one side with the result shown in Fig. 3.6.

4.10 Use $E = \sigma(1 - a/r)/2\varepsilon_0$ from section 2.6. Hence if P is 1/100 mm from a surface, all but 1% of $\sigma/2\varepsilon_0$ is produced by charge less than 1 mm from P.

4.11 The charges near P produce a field $\sigma/2\varepsilon_0$ directed both inward and outward. Since the field inside the conductor is zero, the rest of the charges must produce $\sigma/2\varepsilon_0$ directed outwards at P, thus making the total external field σ/ε_0.

4.12 $E_x = p(2x^2 - y^2)/4\pi\varepsilon_0 r^5$; $E_y = 3pxy/4\pi\varepsilon_0 r^5$ where $r = (x^2 + y^2)^{1/2}$.

4.17 $(2\sqrt{2} - 1)Q^2/32\pi\varepsilon_0 a^2$ towards the intersection of the planes.

4.19 $2\pi(l/(g + Q^2/16\pi\varepsilon_0 a^2 m))^{1/2}$.

4.20 $[(a+d)/(a-d)]^3$.

4.21 $5{\cdot}27 \times 10^{-9}$ cm.

4.22 Potential energy $= \frac{1}{2}\mu x^2$.

CHAPTER 5

5.1 1 pF = 0·9 e.s.u. of capacitance.

5.2 Calculation of V at the midpoint of the axis of the cylinder yields (*a*) uniformly distributed charge, $C = 4\pi\varepsilon_0 l/\log_e\left(\dfrac{l + (a^2 + l^2)^{1/2}}{a}\right)$, (*b*) charge only on corners, $C = 4\pi\varepsilon_0(a^2 + l^2)^{1/2}$. As $l \to \infty$, any end effects produce negligible changes at the centre and (*a*) becomes more accurate. For large l, this tends to $4\pi\varepsilon_0 l/\log_e(2l/a)$.

5.3 (*a*) About 1 pF; (*b*) 760 pF assuming the second formula of problem 5.2; (*c*) 0·8 pF; (*d*) 800 μF.

5.4 Uniform distribution gives $C_u = \varepsilon_0 2a/x$ per unit length; edge distribution gives $C_e = 2\pi\varepsilon_0/\log_e(1 + x^2/a^2)$ per unit length.

5.5 Ideal two-dimensional condenser has capacitance given by C_u of problem 5.4. $C_u/C_e = [\log_e(1 + p^2)]/\pi p$ where $p = x/a$ from problem 5.4 and this is < 1 for all p. Since the real C lies between C_u and C_e, edge effects increase C.

5.6 In series, $V_A = 40$ V, $V_B = 100$ V, $Q = 20$ μC on both; in parallel, $Q_A = 70$ μC, $Q_B = 28$ μC, $V = 140$ V on both.

5.8 Treat the condenser as two in parallel (common V) of capacitances $4\pi\varepsilon_0 ab/(b - a)$ and $4\pi\varepsilon_0 b$. Hence charges are in the ratio $a/(b - a)$.

5.9 (*a*) 159 pF; (*b*) 0·21 μF; (*c*) 0·0053 μF.

5.10 0·0106 μF.

5.11 In series: stored $1{\cdot}4 \times 10^{-3}$ J, supplied $2{\cdot}4 \times 10^{-3}$ J; in parallel: stored $6{\cdot}86 \times 10^{-3}$ J, supplied $1{\cdot}37 \times 10^{-2}$ J.

5.12 48 V; increased 4 times.

5.16 $\varepsilon_0 A V^2/4x$.

5.17 $F = \frac{1}{2}V^2\,\mathrm{d}C/\mathrm{d}x = \pi\varepsilon_0 V^2/\log_e(b/a)$.

5.19 (*a*) The electroscope has a charge opposite to that of the body and thus a potential of opposite sign. As the body approaches, the potential decreases to zero and then increases. (*b*) The charge on the suspended body is of the same sign as but smaller than that of the object. The effect of the opposite induced charge predominates when the bodies are close.

5.20 3 V.

5.21 4 V.

CHAPTER 6

6.1 (*a*) 162,000 C; (*b*) $1{\cdot}944 \times 10^6$ J; (*c*) $22\frac{1}{2}$ hr; (*d*) 20 min.

6.2 $(40 - 0{\cdot}71n)\Omega$; 360 W delivered of which 234 W are dissipated in resistances; 42·6 V NOT 42 V.

6.4 $7R/12$; $5R/6$.

6.5 $3R$, $2{\cdot}75R$ etc. Since an infinite ladder is still infinite when one section is omitted, the resistance looking in at AA is the same as that at the terminals: hence the input resistance is $(\sqrt{3}+1)R$.

6.6 2/3 Ω. The ends of β are at the same potential.

6.7 See Fig. A.1.

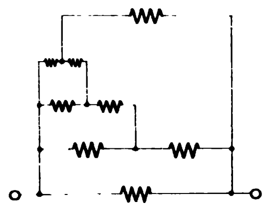


Fig. A.1. Answer to problem 6.7.

6.8 25 Ω, 50 V. (*a*) in series with 100 Ω rheostat, (*b*) in parallel.

6.9 $\mathscr{E}/2R$ in all but two opposite arms: zero in these. Use superposition.

6.10 19/24.

6.12 1/4,880 A.

6.13 The formulae of section 6.3 and Kirchhoff's laws still apply, but not the theorems of section 6.4 which depend on the linearity of equations such as (6.19).

6.14 The *minimum heat theorem* is general and can be deduced from Kirchhoff's laws.

6.15 $RC_1C_2/(C_1+C_2)$.

6.16 $(b-a)/4\pi\sigma ab$. The capacitance of the same system with a dielectric instead of a conductor is $4\pi\varepsilon_r\varepsilon_0 ab/(b-a)$ and thus $CR=\varepsilon_r\varepsilon_0/\sigma$.

6.17 Resistance $1/\pi\sigma(b^2-a^2)$; conductance $2\pi\sigma/\log_e(b/a)$. For radial currents the resistance decreases with increasing length and to quote a value of resistance *per* unit length would be incorrect.

6.18 $\varepsilon_r\varepsilon_0/\sigma$.

6.19 1·25 miles from A.

6.20 $R-R_1{}^2/(R_1+R_L)$. The input resistance is variable for small loads.

6.21 Voltmeter across R only: $1/R=1/R_{app}-1/R_V$, use when $R_V \gg R$; Voltmeter across R and ammeter: $R=R_{app}-R_A$, use when $R \gg R_A$.

6.22 Insert two resistances R_1 and R_2 in series with the driving cell and wire, R_2 being adjacent to the wire at its left-hand end. A current of exactly 50 mA can be obtained by making R_2 a standard 20 Ω resistor and adjusting R_1 (about 18 Ω) until the standard cell gives a balance across R_2 and exactly 18·3 cm of the wire.

6.23 $I_G=(\mathscr{E}_2-R_1\mathscr{E}_1/R)/(r+R_1)-R_1{}^2/R$; $\mathscr{E}_2=R_1\mathscr{E}_1/R$.

CHAPTER 7

7.2 $10^{-3}\pi/4$ A-m^2; $3\times10^{-9}\pi^2/8$ N.

7.6 $\mu_0 ma^2/2(x^2+a^2)^{3/2}$ along the axis.

7.7 For $r<a$: $-2\mu_0 M/3$. For $r>a$: as if all dipoles were at the centre, i.e. a total moment $4\pi a^3M/3$ at the centre.

CHAPTER 8

8.3 A field is *rate of change* of potential.

8.4 If x is the distance from the axis: $\mu_0 Ix/2\pi a^2$ for $x\leqslant a$; $\mu_0 I/2\pi x$ for $a\leqslant x\leqslant b$; $\mu_0 I(c^2/x-x)/2\pi(c^2-b^2)$ for $b\leqslant x\leqslant c$; 0 for $x\geqslant c$.

8.5 1 Wb/m$^2=10^4$ gauss.

8.6 (*a*) $2\sqrt{2}\mu_0 I/\pi a$, (*b*) $4\mu_0 Ia^2/\pi(a^2+4x^2)(2a^2+4x^2)^{1/2}$, (*c*) $\mu_0 nI\tan(\pi/n)/2\pi a$ all perpendicular to the plane of the coil.

8.7 $\mu_0 I(\sin^2\theta)/4a$.

8.8 $\frac{1}{2}\mu_0 J_s$ parallel to the sheet, perpendicular to J_s and in opposite directions on opposite sides of the sheet.

8.9 $\sin^{-1}(IB_0/mg)$.

8.11 IBa.

8.12 $\pi/50$ N of attraction.

8.13 0·000043 N of attraction.

8.14 $\mu_0 Q\omega/4\pi a$.

8.15 $\frac{1}{2}\mu_0\sigma\omega a$.

CHAPTER 9

9.2 0·8 mV.

9.3 $\pi a^2 B(\sin\delta)/R$; about 113 μC.

9.4 $\pi a^2\omega B_0(\cos\omega t)/R$.

9.5 Current is $\sigma\omega B_0 ba^2(\cos\omega t)/4$, mean power loss is $\pi\sigma\omega^2 B_0{}^2 ba^4/8$. The flux density would be considerably changed by the induced currents, but the expression shows the importance of reducing σ, ω and a to cut down eddy current losses.

9.6 $I=(\mathscr{E}-\Phi\omega/2\pi)/R$. From this, $\mathscr{E}I=\Phi\omega I/2\pi R+RI^2$ and the first term on the right is the mechanical power output of the motor.

9.9 $I=(\mathscr{E}-\Phi\omega\sin\omega t)/R$; power $=\Phi I\sin\omega t$.

9.10 95·5 V; 86%.

9.12 1 H $=10^9$ e.m.u. of inductance.

9.13 320 μH.

9.14 Mutual inductance $=4$ mH; e.m.f. $=2$ mV.

9.15 $I_1 I_2\,\partial M/\partial t$.

9.16 $L_1/4L_2$; $\frac{1}{4}(L_1/L_2)^{1/2}$.

9.17 $\mu_0 a\log_e((b+d)/d)/2\pi$.

9.18 $\mu_0 abI^2/2\pi d(b+d)$.

CHAPTER 10

10.1 Express $\sin^2\omega t$ as $\frac{1}{2}-\frac{1}{2}\cos 2\omega t$.

10.3 10^{-4} s. Circuit is critically damped: maximize (10.16).

10.4 10^4 s^{-1}.

10.5 $\omega_N=\omega_0(1-\frac{1}{8}\delta^2)$ while $\Lambda=2\pi\delta(1+\frac{1}{2}\delta^2)$ so that ω_N only differs from ω_0 by a very small quantity and Λ is nearly proportional to δ.

10.7 $1/CR$.

10.8 4·6 μF.

10.9 (*a*) $1/R$ and $1/\omega L$, (*b*) $R/(R^2+\omega^2L^2)$ and $-\omega L/(R^2+\omega^2L^2)$.

10.10 1,000 Ω; 7·3 H.

10.11 For purely resistive $\mathbf{Z}$, Arg $\mathbf{Z}=0$. Hence $(\omega^2LC-1)(R^2-L/C)=0$. Conditions are $R=(L/C)^{1/2}$ and $\omega=1/(LC)^{1/2}$.

10.12 I_{total} and I_C must be equal in *magnitude*. In the vector diagram, the angle between V and I_{total} or I_C is $\tan^{-1}\omega L/R$: hence $\frac{1}{2}I_L/I_C=\omega L/(R^2+\omega^2L^2)^{1/2}$ and since $I_L=V/(R^2+\omega^2L^2)^{1/2}$ and $I_C=\omega CV$, the result follows.

10.14 $(I_2{}^2+\frac{1}{2}I_1{}^2)^{1/2}$.

10.15 $(\frac{1}{2}I_1{}^2+\frac{1}{2}I_2{}^2+I_1I_2\cos\alpha)^{1/2}$.

10.16 1 V; 206 kc/s; 2·3°.

10.21 $M^4/L^2C^2-4M^2R^2/LC>R^4$.

10.25 Note that $(\sqrt{2}\pm 1)^2 = 3\pm 2\sqrt{2}$.

10.26 $e^{\alpha} = 38$ approx. Hence attenuation is $-20 \log_e 38$ or about 116 dB.

10.27 Velocity $= (LC)^{-1/2}$ m/s where L and C are parameters per metre. Hence signal takes $(LC)^{1/2}$ s to travel length of line with total inductance L and total capacitance C. For artificial line, delay is 5 ms per section.

10.28 $60 \log_e 3$ or about 66 Ω; $Z_L = 165$ Ω.

10.30 2·37 μF. Connect in series.

CHAPTER 11

11.1 $B_\theta = \frac{1}{2}\mu_0 \rho v r$; $E_r = \rho r/2\varepsilon_0$. Resultant force on d$Q$ at r from axis is $\rho r\, dQ(1-\varepsilon_0\mu_0 v^2)/2\varepsilon_0$ away from the axis: this is always positive. For neutral currents, no electric force exists and the magnetic force tends to contract the beam, the 'pinch effect'.

11.2 8 mm.

11.3 2·05 m.

11.4 5,600 Mc/s: in the microwave region with wave-length about 5·4 cm.

11.5 4×10^5 m/s; about 840 eV.

11.7 0·999987; 0·43.

11.10 11,200 Mc/s; about 5,600 Mc/s.

CHAPTER 12

12.2 $4{\cdot}34\times 10^{-3}$ m²/V-s.

12.3 $J = n_p e_p(v_p - v_n) = \rho_p\, \delta v$. Both ρ_p and δv are independent of the observer.

12.4 Can show that $J_O = J_P - v\rho$. Only if $\rho = 0$ does $J_O = J_P$ for all v.

12.5 λ is about 1 Å at room temperature.

12.7 0·025 eV for $T = 300$°K.

CHAPTER 13

13.1 0·015 μC; 0·005 μC and $16\frac{2}{3}$ V.

13.2 0·009 μC and 0·006 μC; 30 V.

13.3 $P \cos\theta$ at a point whose radius makes θ with P. For the second part, see section 13.8.

13.4 2:3.

13.5 $4\pi\varepsilon_0/3 \log_e 2$.

13.7 (*a*) $-\varepsilon_0(\varepsilon_r - 1)AtV^2/2\varepsilon_r x^2$, (*b*) $+\varepsilon_0(\varepsilon_r - 1)AtV^2/2\varepsilon_r x^2$.

13.8 $\pi\varepsilon_r\varepsilon_0 a^2 K^2 \log_e (b/a)$.

13.9 $CV^2(\varepsilon_r - 1)/2a$.

13.11 Use equation (4.12).

CHAPTER 14

14.1 (*a*) increases 10^4 times, (*b*) decreases 100 times, (*c*) increases 100 times.

14.2 $J_{sc} = 150$ A/m; $J_{sM\ iron}$ = nearly 150,000 A/m; $J_{sM\ Cu} = 0{\cdot}0015$ A/m.

14.3 1·2 H; 0·30 Wb/m^2; $749{,}500/\pi$ A/m.

14.5 $500/\pi$ A/m.

14.6 $\mu_0(\mu_r + 1)(\log_e 2)/2\pi$.

14.9 0·0433 Wb/m^2.

14.10 $2\pi \times 10^{-4}$ Wb.

CHAPTER 15

15.7 30 V/m.

15.8 $10^{-11}/4\pi$ N.

15.9 $VI/2\pi rl$.

CHAPTER 16

16.1 2:1.

16.3 I_0/π; $2I_0/\pi$.

INDEX

INDEX